PAPER • OPEN ACCESS

Identification of logical patterns for classification of EEE in space application

To cite this article: I S Masich and V I Orlov 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **537** 042073

IOP Conf. Series: Materials Science and Engineering **537** (2019) 042073 doi:10.1088/1757-899X/537/4/042073

Identification of logical patterns for classification of EEE in space application

I S Masich[1] and V I Orlov[2]

[1] Department of System Analysis and Operation Research, Reshetnev Siberian State University of Science and Technology, Krasnoyarsk, Russia
[2] TTC – NPO PM, Zheleznogorsk, Krasnoyarskiy kray, Russia

E-mail: i-masich@yandex.ru

Abstract. The study deals with evaluation of reliability of onboard equipment on the basis of the additional screening tests and the estimated factory tests. In addition to using a statistical approach to solve this problem, the interest is a method for estimating the reliability of onboard equipment using machine learning and data mining methods. Due to the specifics of the task, it is proposed to use classification algorithms based on rules (logical classification algorithms).

1. The problem of classification of EEE

This paper is devoted to the study of possibilities of solving the problem of predicting the reliability of the electronic component base subjected to additional screening tests.

The system has been developed earlier that allows identifying homogeneous production batches in a mixed batch of electronic, electrical and electromechanical (EEE) components for space use [1]. The system is based on the use of the greedy heuristic algorithm. The system does not require additional tests: the data of additional screening tests and additional non-destructive testing are sufficient to identify homogeneous production batches in the mixed batch.

The initial data for the analysis in solving the problem are the results of test actions on the EEE for controlling the current-voltage characteristics of the input and output circuits of the microcircuits. The data represent a table which shows the consequences of various electrical effects on the elements of a set of similar EEE. It is assumed that components with different performance characteristics (in fact, different EEE batches) will have differences in the obtained test results. This assumption makes it possible to use data analysis methods to implement the required EEE grouping.

The problem of identifying homogeneous EEE batches can be viewed as the problem of automatic grouping of objects or the location problem [2]. But the introduction of grouping algorithms into the process of testing, which use the entire set of attributes and which are a “black box”, is difficult. To include the stage of EEE classification in the test program, we need a method for checking the conformity of a component to the class with the help of logical decision rules. In the conditions of space production, the classification method requires not only the accuracy of the solution of the problem, but at the same time the evidence of the approach used and the interpretability of the decision formulated by it.

In this work we investigate the problem of improving the efficiency of classification through the formation of informative patterns based on logical principles of construction, and we develop procedures to improve the interpretability of the classifier based on a small number of rules in it.

IOP Conf. Series: Materials Science and Engineering **537** (2019) 042073 doi:10.1088/1757-899X/537/4/042073

The construction of such classifiers can be based on various methods, among which the most promising for this problem are the methods of logical classification [3], which have an advantage in high interpretability of the classification results. The interpretability of the results of logical classification in the conditions of space production means the possibility of developing stricter standards for EEE parameters.

The most well-known and used algorithms for constructing rules (lists of rules or decision lists) are "separate and conquer" algorithms [4]. It is based on the principle of exclusion of observations already covered (by any of the rules that have been formed) from further consideration (when building other rules). Such algorithms work quickly, and the result is a list of rules.

But such algorithms have flaws that are significant for the problem under consideration. First, the obtained rules (with the exception of the first one) are not optimal by any criterion, since they are based only on the (remaining) part of the training set. Secondly, there may be too little information to make a decision about the classification, since the decision is made only on the basis of the execution of only one rule (or the failure of all), and not on the basis of the "voting" rules.

This paper discusses the possibility of using a more advanced method of identifying patterns in these data and using them to make decisions about belonging to reliability classes, which is based on combinatorics and optimization and is called "logical analysis of data" [5].

The method of logical analysis of data has a number of exceptional advantages over the methods of the "separate and conquer" type. First, all obtained rules can be optimal according to the criterion used (simplicity, selectivity or evidence, as well as their possible combinations). Secondly, the classifier does not just divide the areas of classes, but builds approximation of areas by a set of rules. To assess the strength of the separation of these areas can be used the concept of the margin. Thirdly, using the voting rules, one can assess the reliability of belonging to each class. In addition, there are many other advantages: the ability to work with different types of attributes, the admissibility of the presence of missing values in the data, etc.

This method has only one drawback, it is a high complexity. But the development and use of combinatorial optimization algorithms that are optimal for a class of problems make it possible to use this approach in a reasonable time, while allowing to extract maximum information from the data.

Thus, the paper solves the problem of estimating the reliability of onboard equipment from the results of additional screening tests of the electronic component base in order to further predict reliability indicators.

2. Logical analysis of data

We confine ourselves to the problem of recognition of objects described by binary attributes and divided into two classes

$$K = K^+ \cup K^- \subset B_2^n,$$

where $B_2 = \{0,1\}$, $B_2^n = B_2 \times B_2 \times \cdots \times B_2$.

The classes do not overlap:

$$K^+ \cap K^- = \varnothing.$$

The observation $X \in K$ is described by a binary vector $X = (x_1, x_2, \ldots, x_n)$ and can be represented as a point in the hypercube of the space of binary attributes B_2^n. Observations of the class K^+ will be called positive points of sample *K*, while observations of the class K^- will be called negative points of the sample.

Let us consider a subset of points from B_2^n, in which some variables are fixed and identical, and the rest variables take an arbitrary value:

$$T = \{x \in B_2^n | x_i = 1 \text{ for } \forall i \in A$$
$$\text{and } x_j = 0 \text{ for } \forall j \in B\}$$

for some subsets $A,B \subseteq \{1,2,\ldots,n\}$, $A \cap B = \varnothing$. This set can also be defined as a Boolean function that takes a true value for the elements of the set:

$$t(x) = (\underset{i\in A}{\wedge} x_i) \wedge (\underset{j\in B}{\wedge} \bar{x}_j)$$

The *pattern P* (or *rule*) in this case is a term that covers at least one observation of a certain class and does not cover an observation of another class [6]. That is, the pattern corresponds to a subcube, which has a non-empty intersection with one of the sets (K^+ or K^-) and an empty intersection with another set (K^- or K^+, respectively). The pattern *P*, which does not intersect with K^-, will be called positive, and the pattern *P'*, which does not intersect with K^+, will be called negative.

3. Search for optimal patterns

Let us select some observation $a \in K^+$. Denote by P^a the pattern covering the observation *a*. Those attributes that are fixed in P^a are equal to the corresponding values of the attributes of the object *a*.

To define the patterns P^a, we introduce the binary variables $Y = (y_1, y_2, \ldots, y_n)$:

$$y_j = \begin{cases} 1, \; i\text{-th attribute is fixed in } P^a, \\ 0, \; \text{in another case.} \end{cases}$$

Some point $b \in K^+$ is covered by the pattern P^a only if $y_i = 0$ for all *i* for which $b_i \neq a_i$. On the other hand, some point $c \in K^-$ will not be covered by the pattern P^a if $y_i = 1$ for at least one variable *i* for which $c_i \neq a_i$.

The condition which indicates that the positive pattern should not contain any point of K^- requires that for each observation $c \in K^-$ the variable y_j takes the value 1 for at least one *j*, for which $c_j \neq a_j$ [7]:

$$\sum_{\substack{j=1 \\ c_j \neq a_j}}^{t} y_j \geq 1 \text{ for any } c \in K^-.$$

Strengthening the constraint to increase error tolerance is done by replacing the number 1 on the right side of the inequality by a positive integer *d*.

On the other hand, a positive observation $b \in K^+$ will be included in the considered subcube, when the variable y_j takes the value 0 for all indices *j* for which $b_j \neq a_j$. Thus, the number of positive observations covered by the *a*-pattern can be calculated as

$$\sum_{b\in K^+} \prod_{\substack{j=1 \\ b_j \neq a_j}}^{t} (1 - y_j).$$

Thus, the task of finding the maximum pattern can be written as a task of finding such values $Y = (y_1, y_2, \ldots, y_n)$ for which the resulting pattern P^a covers as many points $b \in K^+$ as possible and does not cover points $c \in K^-$ [7]:

$$\sum_{b\in K^+} \prod_{\substack{i=1 \\ b_i \neq a_i}}^{n} (1 - y_i) \to \max,$$

$$\sum_{\substack{i=1 \\ c_i \neq a_i}}^{n} y_i \geq 1 \text{ for all } c \in K^{-}.$$

This problem is a conditional pseudo-Boolean optimization problem, that is, an optimization problem in which the objective function and the functions on the left side of the constraint are pseudo-Boolean functions that are real functions of Boolean variables. The objective function and the constraint function in this problem are unimodal monotone pseudo-Boolean functions.

The problem of finding the maximum negative patterns is formulated similarly.

The most valuable are the patterns that have the greatest coverage. The more coverage, the better the pattern displays the image of the class.

The described problem is solved using the greedy algorithm [8]. A more perfect approach is to apply an exact algorithm based on the branch and bound method and search among the boundary points of an admissible region [9].

4. Patterns in batches of EEE

The approach was tested on the mixed batch of the chip 1526LE5.

Examples of the obtained patterns:

Class a (196 components):

(TEST_34 < 5) ∧ (TEST_127 < 4.76) => class a (100%)

(TEST_35 < 5) ∧ (TEST_37 < 5) ∧ (TEST_127 < 4.76) => class a (100%)

Class b (218 components):

(TEST_66 ≥ 0.682) ∧ (TEST_122 ≥ 4.98) => class b (97.2%)

(TEST_66 ≥ 0.682) => class b (97.7%)

Class c (205 components):

(TEST_34 ≥ 5) ∧ (TEST_58 < 0.93) => class c (99%)

Also, the most significant patterns have been identified:

Class a:

Test 127	(97.4 %)
Test 35	(50.5 %)
Test 37	(38.8 %)
Test 34	(31.1 %)

Class b:

Test 66	(96.8 %)
Test 122	(79.8 %)

Class c:

Test 58	(97.1 %)
Test 34	(96.6 %)

5. Conclusions

As a result of the development, a method for assessing the reliability of onboard equipment based on the results of additional screening tests of the electronic component base and factory tests of onboard equipment was proposed.

The basis of the developed algorithmic support is an approach to the identification and use of patterns, called logical analysis of data. In the course of research, this approach has been improved. Optimization models have been constructed that allow finding patterns of various types in the data. The use of different types of patterns can improve decision support in the classification. Thus, the use of strong primary patterns reduces the number of unclassified products, and also reduces the number of conditions in the rules (the number of tests). And the use of strong spanned patterns reduces the classification error.

The results of experiments to identify patterns and build rules for the classification of EEE show the prospects of the studied approach, aimed at strengthening decision support in the classification of EEE.

The conducted studies on the use of classification methods based on rules show that the use of a small number of logical rules (from 1 to 4) is sufficient for recognizing a group of radio components. These rules are a comparison of the value of an attribute with the cut point determined in the process of building rules. Thus, it is sufficient to use a very small number of attributes (selected in a certain way in the process of building rules from the entire set of attributes – test results) for successful classification of EEE.

Thus, as a result of the work, the algorithms for identifying patterns in the data obtained from the results of additional screening tests were investigated. If used in conjunction with the results of destructive physical analysis, the proposed approach can be applied to predict the reliability indicators of the electronic component base.

Acknowledgments
Results were obtained in the framework of the state task 2.5527.2017/8.9 of the Ministry of Education and Science of the Russian Federation.

References
[1] Kazakovtsev L A, Orlov V I, Protsenko V V and Fedosov V V 2016 *Development of algorithmic support for analyzing the homogeneity of EEE batches to complete spacecraft electronic equipment* (Krasnoyarsk: Siberian state aerospace university) p 192
[2] Kazakovtsev L A, Orlov V I, Stupina A A and Masich I S 2014 The problem of classifying the electronic component base *Vestnik SibGAU* **4(56)** 55-61
[3] Boros E, Hammer P L, Ibaraki T, Kogan A, Mayoraz E and Muchnik I 2000 An implementation of logical analysis of data *IEEE Transactions on Knowledge and Data Engineering* **12(2)** 292-306
[4] Furnkranz J 1999 Separate-and-Conquer Rule Learning *Artificial Intelligence Review* **13** 3-54
[5] Hammer P L and Bonates T O 2006 Logical analysis of data - An overview: From combinatorial optimization to medical applications *Annals of Operations Research* **148**
[6] Hammer P L 1986 The Logic of Cause-effect Relationships *Lecture at the International Conference on Multi-Attribute Decision via Operations Research-based Expert Systems* (Passau, Germany : Universitat Passau)
[7] Bonates T O, Hammer P L and Kogan A 2008 Maximum Patterns in Datasets *Discrete Applied Mathematics* **156(6)** 846-61
[8] Antamoshkin A N, Masich I S and Kuzmich R I 2015 Heuristics and criteria for constructing logical patterns in data *IOP Conf. Ser.: Mater. Sci. Eng.* **94** 012003
[9] Kazakovtsev L A and Masich I S 2018 A Branch-and-Bound Algorithm for a Pseudo-Boolean Optimization Problem with Black-Box Functions *Facta Universitatis. Ser. Math. Inform.* **33(2)** 337–60