Technique of Cluster Analysis Application in Semiconductor Elements Classification for Improving Product Quality and Reliability

R. O. Mishanov

Department of Design and Technology of Electronic Systems and Devices,
Samara National Research University named after academician S. P. Korolev (Samara University),
Samara, Russian Federation
E-mail: mishanov91@bk.ru

Received: 14.11.2016

Abstract. The text deals with the information about application of cluster analysis for the purpose of semiconductor elements classification. The technique of semiconductor elements classification on groups based on application of k-means clustering is offered. Using the offered technique the research of CMOS chips selection was conducted. A time delay on the leading edge of the signal and a critical voltage supply were used as informative parameters of the chips. The recommendations about application of this technique in case of large number of informative parameters are offered.

Keywords: cluster analysis, k-means clustering, cluster group, informative parameter, Euclidean distance, dendrogram

INTRODUCTION

Nowadays a problem of improving quality and reliability of the electronic devices relates to the priority areas of the technological development. The most actual aspect is looking for ways to improve quality of the electronic components to achieve an acceptable level of product reliability with minimum time and resource spends.

Predicting reliability is constantly being studied and widely distributed in electronics. At the same time, the main attention is paid to the analysis of elements failures, physics of failures application in electronics and statistical failure data. In [1] there is the information about multi-state and continuous-state system reliability with degradation analysis, maintenance models for large repairable systems. Authors [2] present the results of reliability analysis of MOS chips, bipolar transistors and diodes in stationary and mobile systems. At the same time, methods of predicting failures of complex systems are being developed. Authors [3] describe program HIRAP Honeywell (Honeywell In-Service Reliability Assessment Program), which takes into account design failure rates, manufacturing process failure rates and other causes for equipment removal.

A large number of works are devoted to the development of techniques for applying mathematical methods and finding forecasting models. The use of these methods is justified in terms of financial and labor costs to ensure the forecasting process, as well as in terms of the results, that satisfy manufacturers and consumers. Statistical forecasting methods allow obtaining acceptable accuracy of short and medium-term forecasts. By use of statistical methods the long-term forecast accuracy, which satisfy the developers, is difficult to be obtained due to the necessity of large bulk of data processing, which entails the use of many resources. Also, there is no necessity in obtaining long-term forecasts because in many cases the operating term less than the forecast time. Authors [4] offer Monte Carlo simulation using “Excel” spreadsheet for forecasting reliability of a complex system.

The application of the theory of pattern recognition is of special interest. Authors [5] describe the developed system of principles and criteria of determining the relationship of a specified electronic product with one of the preliminary identified product classes. Authors [6] offer the forecasting models for the samples of CMOS chips and Zener diode by an extrapolation method. Thus, research into the application of the theory of pattern recognition to predict the reliability of electronic devices is relevant at this stage in the development of science and technology.

The cluster analysis is related to the theory of pattern recognition. Cluster analysis is a system of data processing algorithms for the distribution of the objects in the group of clusters (cluster groups), which are instantiated by homogeneous objects.

Use of cluster analysis in the practical solution of data analysis problems is often characterized by a known beforehand number of cluster groups. In this case, k-means clustering method got widespread. It involves a construction of *k* different cluster groups, which are located at the greatest possible distance from one another. Initially *k* randomly selected cluster groups are formed, then each object belonging to each group is examined. The criterion for such relationship is to minimize the variability within the cluster and to maximize variability between clusters.

This paper offers the technique of the microchips classification based on the k-means clustering method. The technique is comprised of 3 steps:

– Step 1. Analysis of the initial data. Measurement scale settings. Justification and a choice of proximity measures.

– Step 2. Conducting cluster analysis using method of hierarchical classification. Dendrogram analysis and clusters quantification.

– Step 3. Conducting cluster analysis by k-means clustering. Analysis of the results.

INITIAL DATA

CMOS chips 765LN2 sample was investigated. The integrated circuits are represented by six logic elements NAND. The sample includes 50 chips. Time delay on the leading edge of the signal x_1 (t^+_p, μs) and critical voltage supply x_2 (V_{sc}, V) are selected as the informative parameters. Table 1 presents the initial data for cluster analysis.

PROBLEM FORMULATION

The research objective is to split elements into cluster groups (classes). The elements in one group are characterized by a similar condition.

PROBLEM SOLUTION

Task solution by the cluster analysis performed using the program package STATISTICA 10.

The first step of the solution is to set a measurement scale, as cluster groups are characterized by assessing the distance between elements. But as the measurement scale of values x_1 and x_2 is different (x_1 is measured in microseconds, x_2 – in volts), the parameter values must be centered, i.e. lead to the scale when average value of the parameters equals 0 and standard deviation equals 1.

Table 1. The initial data for cluster analysis

The element number	X_1	X_2	**The element number**	X_1	X_2	**The element number**	X_1	X_2
1	4.30	1.30	19	3.80	1.20	37	3.80	1.40
2	7.20	2.90	20	3.70	1.20	38	6.40	2.50
3	3.20	1.10	21	4.40	1.30	39	6.00	2.50
4	6.60	2.10	22	7.10	2.60	40	6.90	2.81
5	5.30	1.72	23	5.10	1.60	41	7.10	1.74
6	4.70	1.60	24	5.00	1.50	42	5.00	1.72
7	6.70	1.80	25	15.60	3.50	43	7.30	2.90
8	6.20	1.73	26	5.00	2.80	44	8.10	2.80
9	6.60	2.40	27	4.40	1.71	45	5.20	2.30
10	3.90	1.30	28	4.50	1.80	46	7.10	2.81
11	4.50	1.40	29	3.00	1.00	47	10.80	3.20
12	4.30	1.40	30	4.20	1.56	48	3.50	1.50
13	4.60	1.50	31	4.90	1.60	49	4.00	1.63
14	5.80	1.70	32	7.50	2.80	50	5.10	1.90
15	9.20	2.90	33	4.50	1.62			
16	6.50	2.60	34	7.80	2.78			
17	7.00	2.80	35	8.90	2.96			
18	5.20	1.75	36	4.80	1.70			

Cluster analysis involves the use one of the ways to determine the proximity measure, the use of which depends on the final version of the partition of objects into clusters. Selection of the proximity measure depends on the research objectives, the nature of the probability distribution and other parameters. In this example, the ordinary Euclidean Distance is used as the proximity measure, because a sample is taken from a population with normal distribution, and informative parameters x_1 and x_2 are equally important for the classification.

Centered values of the informative parameters are shown in Table 2. Rounding made to the third decimal place.

The Euclidean Distance is calculated for each pair of elements by the formula:

$$\rho_E(a_i, a_j) = \sqrt{(x_{1ci} - x_{1cj})^2 + (x_{2ci} - x_{2cj})^2}, \qquad (1)$$

where a_i and a_j – the i^th^ and j^th^ elements of the sample; x_{1ci} and x_{1cj} – centered characteristic value of parameter x_1 of the i^{th} and j^{th} elements of the sample; x_{2ci} and x_{2cj} – centered characteristic value of parameter x_2 of the i^{th} and j^{th} elements of the sample.

In the program package STATISTICA a cluster analysis using the hierarchical classification was selected, the rule of association “total communication method” was noted, the “Euclidean distance” was selected as a measure of proximity. The results of hierarchical classification are shown in Fig. 1. The vertical dendrogram of the cluster analysis using hierarchical classification is shown in Fig. 2.

Table 2. Centered values of the informative parameters

The element number	X_{1c}	X_{2c}	The element number	X_{1c}	X_{2c}	The element number	X_{1c}	X_{2c}
1	–0.709	–1.097	19	–0.938	–1.249	37	–0.938	–0.944
2	0.621	1.345	20	–0.984	–1.249	38	0.254	0.734
3	–1.214	–1.402	21	–0.663	–1.097	39	0.071	0.734
4	0.346	0.124	22	0.575	0.887	40	0.483	1.207
5	–0.250	–0.456	23	–0.342	–0.639	41	0.575	–0.425
6	–0.526	–0.639	24	–0.388	–0.792	42	–0.388	–0.456
7	0.392	–0.334	25	4.474	2.260	43	0.667	1.345
8	0.162	–0.441	26	–0.388	1.192	44	1.034	1.192
9	0.346	0.582	27	–0.663	–0.471	45	–0.296	0.429
10	–0.893	–1.097	28	–0.617	–0.334	46	0.575	1.207
11	–0.617	–0.944	29	–1.305	–1.555	47	2.272	1.802
12	–0.709	–0.944	30	–0.755	–0.700	48	–1.076	–0.792
13	–0.572	–0.792	31	–0.434	–0.639	49	–0.847	–0.593
14	–0.021	–0.486	32	0.759	1.192	50	–0.342	–0.181
15	1.538	1.345	33	–0.617	–0.609			
16	0.300	0.887	34	0.896	1.162			
17	0.529	1.192	35	1.401	1.436			
18	–0.296	–0.410	36	–0.480	–0.486			

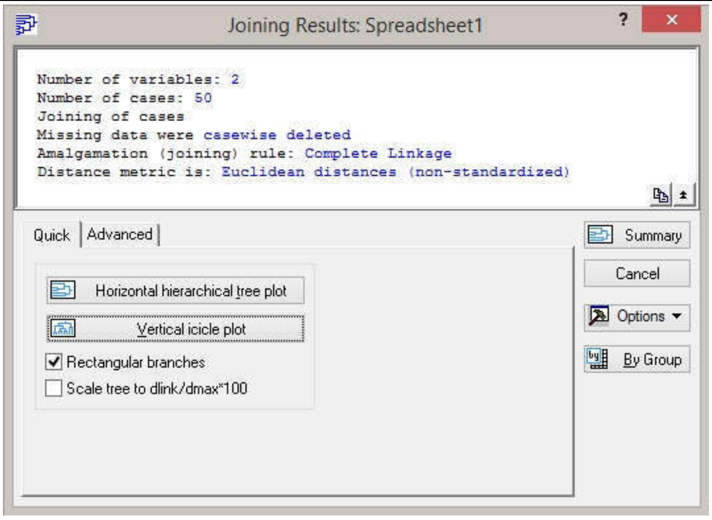

Figure 1. The results of hierarchical classification of clustering analysis

The meaning of the vertical dendrogram is as follows. The shorter a distance between the elements, the more similar in informative parameters they are, therefore they belong to the same cluster. With increasing a distance between the elements the differences become larger. Each node of the dendrogram indicates the union of two or more clusters. The distance, at which the clusters are combined, is indicated on the vertical axis. The horizontal axis shows the number of elements.

Analysis of the dendrogram demonstrates that on a distance of 6.924 the element No.25 combined with a large cluster, which is formed on a distance of 4.9065. It can be concluded that the optimal number of clusters is 2, one of which is formed on a distance of 2.9142, and the second - on a distance of 2.3456. The element No.25 can be considered as the outlier.

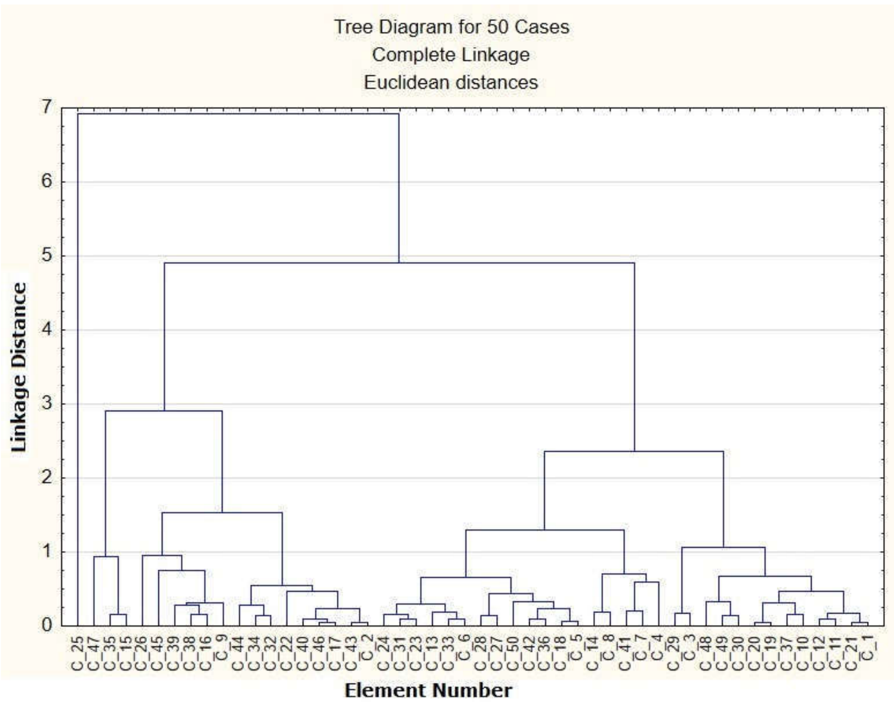

Figure 2. The vertical dendrogram of the cluster analysis using hierarchical classification

The next step is carrying out a cluster analysis of sample using k-means clustering for testing the significance of differences between groups of clusters. Using this method you can specify the initial number of clusters, which are the center of group elements with the most similar parameters. Thus at each iteration object composition of clusters varies. The optimality criterion is a minimization of variability in the cluster and maximization of variability between clusters.

Cluster analysis with clusterization by k-means clustering was carried out in the program package STATISTICA. Based on the analysis of the dendrogram shown in Fig. 2, take the number of clusters equals 2. The number of iterations is standard and equals 10. The results of the cluster analysis by k-means clustering are shown in Fig. 3.

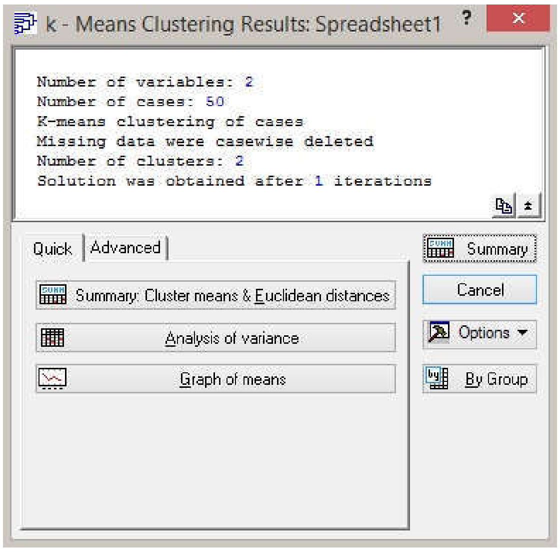

Figure 3. The results of the cluster analysis by k-means clustering

To determine the significance of differences between groups of clusters the dispersion analysis should be carried out. The results are shown in Fig. 4.

	Analysis of Variance (Spreadsheet1)					
Variable	**Between SS**	df	Within SS	df	F	signif. p
X1	23,82950	1	25,17050	48	45,4427	0,000000
X2	40,43696	1	8,56304	48	226,6688	0,000000

Figure 4. The results of the dispersion analysis

The significance level *p* of each informative parameter less than 0.05 (acceptable limit of error level). Thus, a significant difference in both parameters between clusters exists.

It is necessary to define the distance from each element included in the cluster to the cluster center. In the program the button “Summary: Cluster means & Euclidean distances” was selected (see Fig. 5).

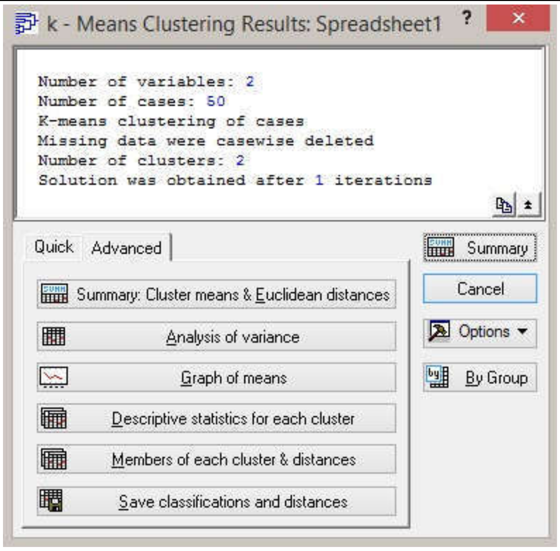

Figure 5. Selection of items for sample analysis

Fig. 6a, 6b show the composition of the clusters and the distances between elements to the clusters center calculated by the k-means clustering. The number of clusters specified in initial data for the k-means clustering, equals 2. The element No.25, which was considered as outlier, is contained in the cluster closest to him (cluster No.2). The distance between it and the center of the cluster is relatively longer than distances between other elements of the cluster and the cluster center.

Fig. 6b shows that the elements No.25, No.47 have a relatively long distance to the cluster center. In the future it will increase the field of standard deviations in the cluster No.2.

For each informative parameter form the table of average values (table is shown in Fig. 7). Fig. 8 shows a plot of means and confidence intervals for each parameter in each cluster.

According to the plot, which is shown in Fig. 8, it demonstrates that the informative parameter x_1 has a high standard deviation in the cluster No.1. It is explained by the influence of the parameter x_1 values of elements No.25 and No.47. To exclude such a case we could set the number of initial clusters to 3. Then elements No.25 and No.47 form another cluster, which can’t be taken into account in further studies. Therefore, the standard deviation field of the parameter x_1 in the cluster No.2 is significantly reduced.

Case No.	Members of Cluster Number 1 (Spreadsheet1) and Distances from Respective Cluster Center Cluster contains 31 cases Distance
C_1	0,302286
C_3	0,685738
C_5	0,269869
C_6	0,047121
C_7	0,709229
C_8	0,530728
C_10	0,373040
C_11	0,178358
C_12	0,207554
C_13	0,065737
C_14	0,398170
C_18	0,270154
C_19	0,477439
C_20	0,497250
C_21	0,291017
C_23	0,147539
C_24	0,124308
C_27	0,186136
C_28	0,267342
C_29	0,808927
C_30	0,151737
C_31	0,088238
C_33	0,086717
C_36	0,159715
C_37	0,328735
C_41	0,813128
C_42	0,205884
C_45	0,819640
C_48	0,383790
C_49	0,230306
C_50	0,395344

(a)

Case No.	Members of Cluster Number 2 (Spreadsheet1) and Distances from Respective Cluster Center Cluster contains 19 cases Distance
C_2	0,230634
C_4	0,817758
C_9	0,551701
C_15	0,484567
C_16	0,451148
C_17	0,251116
C_22	0,285093
C_25	2,658986
C_26	0,898459
C_32	0,092305
C_34	0,013692
C_35	0,419575
C_38	0,531855
C_39	0,644105
C_40	0,284706
C_43	0,205632
C_44	0,111834
C_46	0,220730
C_47	1,086533

(b)

Figure 6. The distances to the cluster center of each element, which is part of: the cluster No.1 (a); the cluster No.2 (b)

Breakdown Table of Descriptive Statistics (Spreadsheet1)
N=50 (No missing data in dep. var. list)

CLUSTER	X1 Means	X1 N	X1 Std.Dev.	X1 Minimum	X1 Maximum	X2 Means	X2 N	X2 Std.Dev.	X2 Minimum	X2 Maximum
1	4,667742	31	0,927501	3,000000	7,10000	1,557419	31	0,263097	1,000000	2,300000
2	7,768421	19	2,283042	5,000000	15,60000	2,771579	19	0,298222	2,100000	3,500000
All Grps	5,846000	50	2,180078	3,000000	15,60000	2,018800	50	0,655329	1,000000	3,500000

Figure 7. Table of average values of each informative parameter

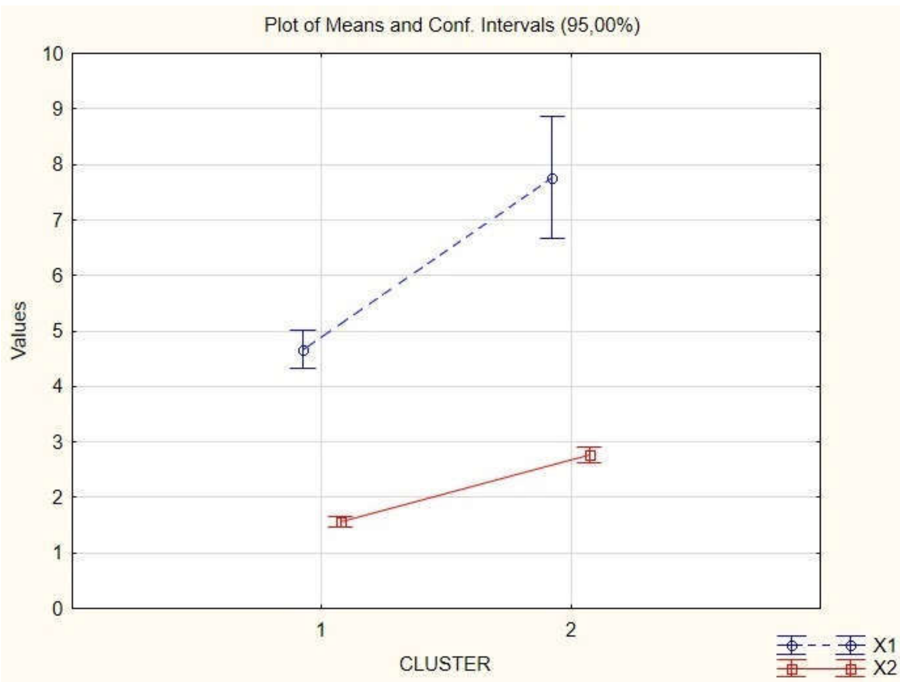

Figure 8. Plot of means and confidence intervals for each parameter in each cluster

CONCLUSION

Analyzing the results it is stated, that the use of cluster analysis for classification of semiconductor elements into classes have meaning when the element classes are well separated. Otherwise, the characteristic values of some elements can increase the standard deviation field that in the future will lead to the accuracy decrease of classification of the element from the party to a particular class.

Increasing number of informative parameters there is a problem to measure the effect of each particular parameter to the selected distance metric. In this case, it is advisable to use the Weighted Euclidean Distance as a distance metric. Thus, the "weight" is attributed to each of informative parameters. Therefore, the influence degree of each parameter on distance metric is taken into account. The difficulty lies in the evaluation of these "weights". One of the best options for the "weights" evaluation of each parameter is the expert evaluation method.

REFERENCES

1. Frenkel, I. B., Karagrigoriou, A., Lisnianski, A., & Kleyner, A. (Eds.) (2014). *Applied reliability engineering and risk analysis: Probabilistic models and statistical inference*. Chichester, UK: John Wiley & Sons Inc. doi:10.1002/9781118701881
2. Campbell, D. S., Hayes, J. A., Jones, J. A., & Schwarzenberger, A. P. (1992). Reliability behavior of electronic components as a function of time. *Quality and Reliability Engineering, 8*(3), 161–166. doi:10.1002/qre.4680080303
3. Johnson, B. G., & Gullo, L. (2000). Improvements in reliability assessment and prediction methodology. *Annual Reliability and Maintainability Symposium. 2000 Proceedings. International Symposium on Product Quality and Integrity (Cat. No.00CH37055)*. Los Angeles, USA. 181–187. doi:10.1109/rams.2000.816304
4. Gedam, S. G., & Beaudet, S. T. (2000). Monte Carlo simulation using Excel(R) spreadsheet for predicting reliability of a complex system. *Annual Reliability and Maintainability Symposium. 2000 Proceedings. International Symposium on Product Quality and Integrity (Cat. No.00CH37055)*. Los Angeles, USA. 188–193. doi:10.1109/rams.2000.816305
5. Georgiev, A., & Georgieva, N. (2002) Reliability prediction of the electronic products. *Proceedings of 37-th International Scientific Conference on Information, Communication and Energy Systems and Technologies – Session EL Electronics (ICEST 2002)*, Nis, Yugoslavia, *1,* 85–88.
6. Mishanov, R., & Piganov, M. (2015). Individual forecasting of quality characteristics by an extrapolation method for the samples of the stabilitrons and the integrated circuits. *13th International Conference The Experience of Designing and Application of CAD Systems in Microelectronics (CADSM)*. Lviv, Ukraine: IEEE. doi:10.1109/cadsm.2015.7230846