

УДК 621.382;51.74

МЕТОДИКА ПРИМЕНЕНИЯ КЛАСТЕРНОГО АНАЛИЗА ДЛЯ КЛАССИФИКАЦИИ ЭЛЕКТРОРАДИОИЗДЕЛИЙ И ПОВЫШЕНИЯ НАДЕЖНОСТИ АППАРАТУРЫ

© 2017 Р.О. Мишанов

Самарский национальный исследовательский университет имени академика С.П. Королёва

Статья поступила в редакцию 14.03.2017

Приведена методика применения кластерного анализа для классификации экспериментальной базы и повышения качества и надёжности радиоэлектронной аппаратуры. Для проведения кластерного анализа использован метод К-средних. Исследования проведены на выборке микросхем 765ЛН2. Было использовано центрирование исходных данных, в качестве которых выступали информативные параметры. С помощью иерархической классификации получена вертикальная дендрограмма кластерного анализа. Проведен дисперсионный анализ результатов, который позволил определить значимость различий между кластерами. Установлено, что кластерный анализ эффективен в том случае, когда классы экземпляров хорошо разделяются.

Ключевые слова: *кластерный анализ, методика, классификация, радиоэлектронная аппаратура, электрорадиоизделие, микросхема, информативный параметр, центрирование, дендрограмма, дисперсионный анализ*

Проблема повышения качества и надёжности электрорадиоизделий (ЭРИ) и радиоэлектронной аппаратуры (РЭА) в современном мире относится к приоритетным направлениям развития техники. Наиболее актуальным на сегодняшний день является поиск способов, позволяющих с минимальными затратами времени и ресурсов обеспечить приемлемый уровень качества и надёжности элементно-компонентной базы (ЭКБ) и аппаратуры. К таким способам можно отнести прогнозирование будущего состояния аппаратуры [1, 2]. С развитием средств вычислительной техники широкое распространение в прогнозировании получили вероятностные методы, использующие математические модели [3-5]. Использование таких методов для прогнозирования показателей качества и надёжности ЭРИ и технических изделий обосновано, как с точки зрения финансовых и трудовых затрат, так и с точки зрения получения результатов, удовлетворяющих изготовителей и потребителей. Вероятностные методы прогнозирования позволяют получить приемлемую точность краткосрочных и среднесрочных прогнозов.

Одним из перспективных направлений в прогнозировании является теория распознавания образов, к которой относится кластерный анализ, представляющий собой систему алгоритмов обработки данных для распределения исследуемых объектов на группы-кластеры, характеризующиеся однородностью объектов [6, 7]. В отличие от дискриминантного анализа и других методов кластерный анализ не предполагает разбиение объектов на заранее известные классы, характеризующиеся заранее определенными параметрами. С одной стороны, если мы не можем изначально задать классы (например, класс годных, негодных ЭРИ), то в таком случае применение кластерного анализа не обосновано. Но с другой стороны, разбиение совокупности

объектов на какое-то конечное число кластеров может дать информацию о «степени сходства» объектов внутри кластера, что при анализе результатов поможет сформулировать выводы.

Цель работы: разработка и апробация методики применения кластерного анализа для классификации ЭРИ и повышения надёжности РЭА.

Методика классификации. В этой работе для классификации ЭРИ с целью прогнозирования состояния изделий с помощью кластерного анализа предлагается следующая методика:

- 1 этап. Анализ исходных данных. Задание масштаба измерений с помощью центрирования значений информативных параметров. Обоснование и выбор меры близости.

- 2 этап. Проведение кластерного анализа с помощью иерархической классификации. Анализ дендрограммы и определение количества кластеров.

- 3 этап. Проведение кластерного анализа методом К-средних. Анализ результатов.

Исходные данные. В ходе работы была исследована выборка микросхем 765ЛН2, выполненных по КМОП-технологии и представляющих собой шесть логических элементов И-НЕ. Объем выборки микросхем составил 50 микросхем. В качестве информативных параметров были выбраны следующие [8]: x_1 – время задержки по переднему фронту сигнала $t_{\text{р}}$, [мкс]; x_2 – критическое питающее напряжение $E_{\text{кр. п}}$, [В]. Исходные данные для проведения кластерного анализа представлены в табл. 1.

Постановка и решение задачи. Целью данного исследования является разбиение экземпляров на группы (классы). Экземпляры, попавшие в одну группу, характеризуются одинаковой вероятностью отказа.

Решение задачи кластерным анализом проводилось с помощью пакета программы STATISTICA 10 [9, 10]. На первом этапе решения был задан масштаб измерений, т.к. кластерные группы характеризуются

оценкой расстояния между экземплярами. Но так как масштаб измерения величин x_1 и x_2 разный, то значения параметров центрировали, т.е. приводили

к шкале, при которой переменные имеют среднее значение 0, а стандартное отклонение 1 [11].

Таблица 1. Исходные данные для проведения кластерного анализа

№ экз.	X ₁	X ₂	№ экз.	X ₁	X ₂	№ экз.	X ₁	X ₂
1	4,3	1,3	19	3,8	1,2	37	3,8	1,4
2	7,2	2,9	20	3,7	1,2	38	6,4	2,5
3	3,2	1,1	21	4,4	1,3	39	6	2,5
4	6,6	2,1	22	7,1	2,6	40	6,9	2,81
5	5,3	1,72	23	5,1	1,6	41	7,1	1,74
6	4,7	1,6	24	5,0	1,5	42	5,0	1,72
7	6,7	1,8	25	15,6	3,5	43	7,3	2,9
8	6,2	1,73	26	5	2,8	44	8,1	2,8
9	6,6	2,4	27	4,4	1,71	45	5,2	2,3
10	3,9	1,3	28	4,5	1,8	46	7,1	2,81
11	4,5	1,4	29	3	1	47	10,8	3,2
12	4,3	1,4	30	4,2	1,56	48	3,5	1,5
13	4,6	1,5	31	4,9	1,6	49	4	1,63
14	5,8	1,7	32	7,5	2,8	50	5,1	1,9
15	9,2	2,9	33	4,5	1,62			
16	6,5	2,6	34	7,8	2,78			
17	7,0	2,8	35	8,9	2,96			
18	5,2	1,75	36	4,8	1,7			

Кластерный анализ предполагает использование одного из способов определения меры близости, в зависимости от применения которого зависит окончательный вариант разбиения объектов на кластеры. Выбор меры близости зависит от целей исследования, характера вероятностного распределения и других параметров. В нашем примере в качестве меры близости использовалось обычное

Евклидово расстояние, т.к. выборка была взята из генеральной совокупности, имеющей нормальное распределение, а информативные признаки x_1 и x_2 одинаково важны для классификации [10]. Центрированные значения информативных параметров приведены в табл. 2. Округление производилось до третьего знака после запятой.

Таблица 2. Центрированные значения информативных параметров

№ экз.	X _{1ц}	X _{2ц}	№ экз.	X _{1ц}	X _{2ц}	№ экз.	X _{1ц}	X _{2ц}
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			

Обычное Евклидово расстояние вычисляется для каждой пары экземпляров по формуле:

$$\rho_E(a_i, a_j) = \sqrt{(x_{1цi} - x_{1цj})^2 + (x_{2цi} - x_{2цj})^2}, \quad (1)$$

где a_i и a_j – i -ый и j -ый экземпляры выборки; $x_{1цi}$ и $x_{1цj}$ – центрированные значения признака x_1 i -ого и j -ого экземпляров выборки соответственно; $x_{2цi}$ и $x_{2цj}$ – центрированные значения признака x_2 i -ого и j -ого экземпляров выборки соответственно [12].

Выясним, формируются ли кластеры, которые могут быть осмыслены [11]. В программе STATISTICA 10 выберем проведение кластерного анализа с помощью иерархической классификации, укажем правило объединения «методом полной связи», а мерой близости – «Евклидово расстояние». Результаты иерархической классификации представлены на рис. 1. Вертикальная дендрограмма проведения кластерного анализа иерархической классификацией представлена на рис. 2. Смысл вертикальной дендрограммы заключается в следующем. Чем меньше «расстояние» между экземплярами, тем они более схожи по информативным параметрам, а значит, относятся к одному кластеру. С увеличением «расстояния» между экземплярами увеличиваются и различия. Каждый узел дендрограммы указывает на объединение двух или более кластеров. «Расстояние», на котором объединены кластеры, указывается по вертикальной оси. По горизонтальной оси показаны номера экземпляров элементов. Анализ дендрограммы наглядно показывает, что при расстоянии 6,924 экземпляр №25 объединяется с крупным кластером, который образуется

при расстоянии 4,9065. Можно сделать вывод, что оптимальное количество кластеров равно 2, первый из которых формируется при расстоянии 2,9142, а второй – при расстоянии 2,3456. Экземпляр 25 можно признать выпадающим значением.

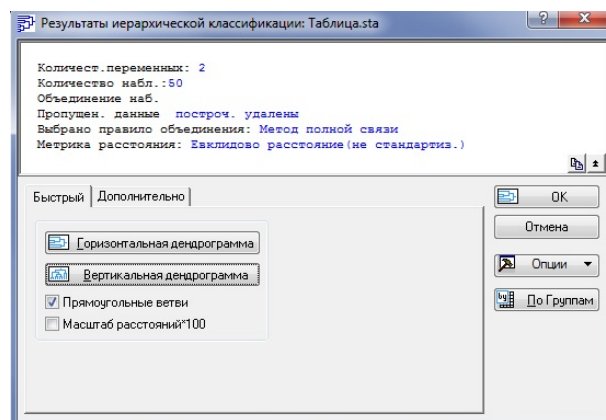


Рис. 1. Результаты кластерного анализа иерархической классификацией

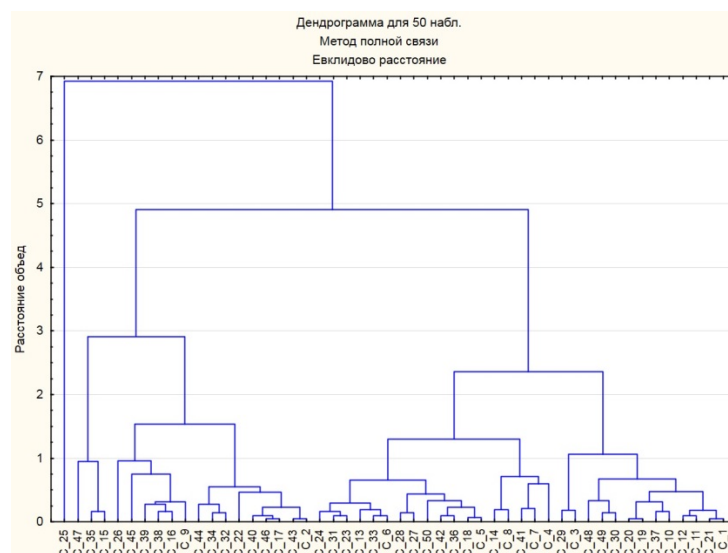


Рис. 2. Вертикальная дендрограмма проведенного кластерного анализа с помощью иерархической классификации

Кластерный анализ. На следующем этапе проведем кластерный анализ выборки элементов с помощью кластеризации методом К-средних для проверки значимости различия между группами-кластерами [11]. При таком методе можно задать изначальное количество кластеров, представляющих собой центры, вокруг которых собираются экземпляры с наиболее близкими параметрами. При этом при каждой итерации решения объектный состав кластеров меняется. Критерием оптимальности является минимизация изменчивости внутри кластера и максимизация изменчивости между кластерами. Проведем кластерный анализ с кластеризацией методом К-средних с помощью программы STATISTICA 10. Исходя из анализа дендрограммы, приведенной на рис. 1, примем в качестве исходных

данных число кластеров, равное 2. Число итераций оставим стандартное, равное 10. Результаты проведения кластерного анализа методом К-средних приведены на рис. 3. Для определения значимости различия между группами-кластерами проведем дисперсионный анализ. Результаты анализа представлены на рис. 4.

Отметим, что уровень значимости p по каждому информативному параметру менее 0,05 (примемлемая граница уровня ошибки). Таким образом, между кластерами существует значимое различие по обоим параметрам. Определим расстояние до центра кластера от каждого элемента, входящего в кластер. Для этого в программе необходимо выбрать пункт «Элементы кластеров и расстояния» (рис. 5).

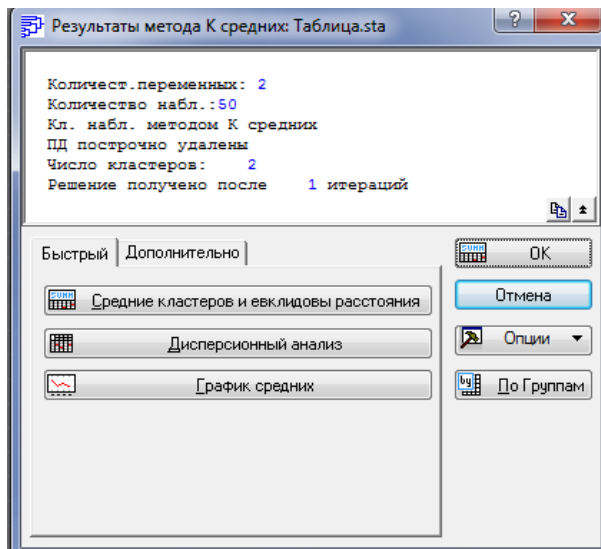


Рис. 3. Результаты проведения кластерного анализа методом К-средних

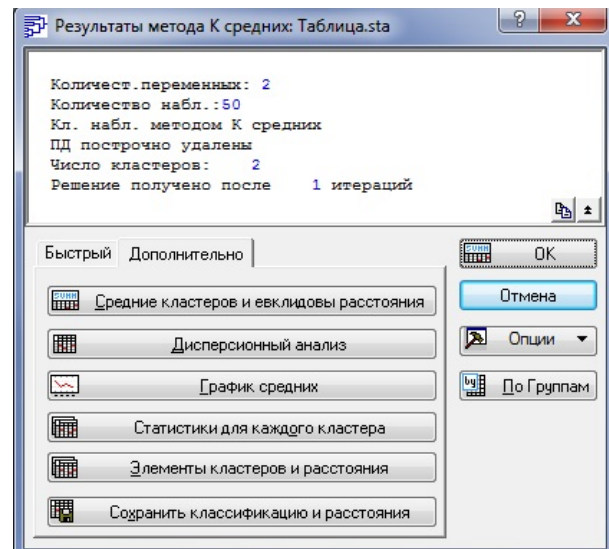


Рис. 5. Выбор пунктов для анализа выборки

перемен.	Дисперсионный анализ (Таблица.sta)					
	Между SS	сс	Внутри SS	сс	F	значим. р
X1	23,82950	1	25,17050	48	45,4427	0,000000
X2	40,43696	1	8,56304	48	226,6688	0,000000

Рис. 4. Результаты дисперсионного анализа

Элементы кластера номер 1 (Таблица.sta) и расстояния до центра кластера. Кластер содержит 31 набл.	
Наблюд.	объедин.
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

а)

Элементы кластера номер 2 (Таблица.sta) и расстояния до центра кластера. Кластер содержит 19 набл.	
Наблюд.	объедин.
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

б)

Рис. 6. Расстояния до центра кластера от каждого экземпляра, входящего в состав группы-кластера: а) элементы кластера 1; б) элементы кластера 2

На рис. 6 показан состав кластеров и расстояния между экземплярами до центра кластеров, вычисленные с помощью метода К-средних. Отметим,

что количество кластеров, заданных в исходных данных для метода К-средних, равнялось двум. Экземпляр №25, который при анализе дендрограммы

иерархической классификации приняли выпадающим, содержится в наиболее близком ему кластере (кластер 2). По этой причине расстояние между ним и центром кластера сравнительно больше расстояний между другими экземплярами кластера и центром этого кластера. По рис. 6б видно, что экземпляры №25 и №47 имеют сравнительно большое расстояние до центра кластера. В дальнейшем это приведет к увеличению поля стандартных отклонений в кластере 2.

Для каждого из параметров сформируем таблицу средних значений (указана на рис. 7). На рис. 8 показан график средних значений и доверительных интервалов (95%) для каждого параметра в каждом

кластере. Из графика, изображенного на рис. 8, видно, что информативный параметр x_1 имеет высокое значение стандартного отклонения в кластере 2, что объясняется влиянием значения параметра x_1 экземпляров №25 и №47. Для исключения такого случая можно в исходных данных к проведению этапа кластерного анализа методом К-средних задаться количеством исходных кластеров, равным 3. Тогда экземпляры №25 и №47 образуют еще один кластер, который можно будет не учитывать в дальнейших исследованиях. Тогда поле стандартного отклонения параметра x_1 в кластере 2 значительно уменьшится.

Итоговая таблица средних (Таблица ста)
N=50 (Нет пропусков в завис. перем.)

Кластер	X1 Среднее	X1 N	X1 Ст.откл.	X1 Минимум	X1 Максим.	X2 Среднее	X2 N	X2 Ст.откл.	X2 Минимум	X2 Максим.
1	4.667742	31	0.927501	3.000000	7.100000	1.557419	31	0.263097	1.000000	2.300000
2	7.768421	19	2.283042	5.000000	15.600000	2.771579	19	0.298222	2.100000	3.500000
Всего	5.846000	50	2.180078	3.000000	15.600000	2.018800	50	0.655329	1.000000	3.500000

Рис. 7. Таблица средних значений для каждого информативного параметра

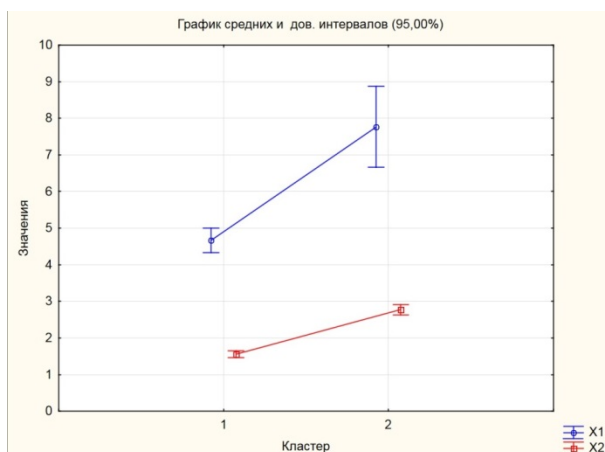


Рис. 8. График средних значений и доверительных интервалов для каждого параметра в каждом кластере

Выводы: предложена методика применения кластерного анализа для классификации ЭРИ и повышения надёжности РЭА. Проведена апробация этой методики на выборке микросхем 765ЛН2. При анализе полученных результатов установлено, что применение кластерного анализа с целью классификации ЭРИ на классы годных и негодных экземпляров эффективен в случае, когда классы экземпляров хорошо разделяются. Иначе значения параметров некоторых экземпляров выборки могут увеличивать поле стандартных отклонений, что в дальнейшем приведёт к снижению точности отнесения элемента из партии к конкретному классу. С увеличением количества рассматриваемых информативных параметров возникает проблема учета влияния каждого конкретного параметра на выбранную меру близости. В таком случае в качестве меры близости целесообразно использовать «взвешенное» Евклидово пространство [12]. В этом случае каждому информативному параметру

приписывается «вес» и, следовательно, учитывается степень влияния каждого параметра на меру близости. Сложность заключается в количественном определении таких «весов». Одним из оптимальных вариантов количественного определения «весов» для каждого параметра является метод экспертных оценок.

СПИСОК ЛИТЕРАТУРЫ:

1. Piganov, M.N. Apparatus diagnostic for non-destructive control chip CMOS-Type / M.N. Piganov, S.V. Tyulevin, E.S. Erantseva, R.O. Mishanov // European science and technology: materials of the VIII international research and practice conference. – Germany, Munich, 2014. P. 398-401.
2. Пиганов, М.Н. Прогнозирование надежности радио-электронных средств / М.Н. Пиганов, С.В. Тюлевин // Научно-технические ведомости СПбГПУ. Серия «Информатика.-Телекоммуникации. Управление». 2009. Вып. 1. С. 175-182.
3. Mishanov, R.O. Individual forecasting of quality characteristics by an extrapolation method for the stabilitrons and the integrated circuits / R.O. Mishanov, M.N. Piganov // The experience of designing and application of CAD systems in Microelectronics (CADSM 2015): Proceeding XIII international conference. – Ukraine, Lviv, 2015. P. 242-244.
4. Piganov, M.N. Individual prognosis of quality indicators of space equipment elements / M.N. Piganov, S.V. Tyulevin, E.S. Erantseva // The experience of designing and application of CAD systems in microelectronics (CADSM 2015): Proceeding XIII international conference. – Ukraine, Lviv, 2015. P. 367-371.
5. Мишанов, Р.О. Разработка прогнозной модели качества полупроводниковых приборов методом экстраполяции / Р.О. Мишанов, М.Н. Пиганов // Известия Самарского научного центра Российской академии наук. 2014. Том 16, №4(3). С. 594-599.
6. Ту Дж. Принципы распознавания образов / Дж.Ту, Р. Гонсалес // Пер. с англ. И.Е. Гуревича: под ред. Ю.И. Журавлёва. – М.: Мир, 1978. 412 с.

7. Мандель, И.Д. Кластерный анализ. – М.: Финансы и статистика, 1988. 176 с.
8. Тюлевин, С.В. К проблеме прогнозирования показателей качества элементов космической аппаратуры / С.В. Тюлевин, М.Н. Пуганов, Е.С. Еранцева // Надежность и качество сложных систем. 2014. № 1(5). С. 9-17.
9. Боровиков, В.П. Прогнозирование в системе STATISTICA в среде Windows / В.П. Боровиков, Г.И. Ивченко // Основы теории и интенсивная практика на компьютере. – М.: Финансы и статистика, 2000. 384 с.
10. Боровиков, В.П. STATISTICA. Искусство анализа данных на компьютере. – СПб.: Питер, 2003. 700 с.
11. Пример использования кластерного анализа STATISTICA в автостраховании [Электронный ресурс] // STATISTICA: Data mining, анализ данных, контроль качества, прогнозирование, обучение, консалтинг [сайт]. [2016]. URL: http://statsoft.ru/solutions/ExamplesBase/branches/detail.php?ELEMENT_ID=1573 (дата обращения: 05.08.2016).
12. Мхитарян, В.С. Эконометрика / В.С. Мхитарян, М.Ю. Архипова, В.П. Сиротин // Учебно-методический комплекс. – М.: Изд. Центр ЕАОИ, 2008. 144 с.

TECHNIQUE OF CLUSTER ANALYSIS APPLICATION FOR ELECTRONIC COMPONENTS CLASSIFICATION AND EQUIPMENT RELIABILITY IMPROVEMENT

© 2017 R.O. Mishanov

Samara National Research University named after acad. S.P. Korolyov

The paper deals with the technique of cluster analysis application for experimental facilities classification and quality and reliability improvement of electronic equipment. K-means clustering method is used for carrying-out of cluster analysis. Studies were conducted on a sample of chips 765LN2. The centering of initial data (informative parameters) was used. By means of hierarchical classification the vertical dendrogram of cluster analysis was obtained. The dispersive analysis of results was carried out. It allowed to determine the importance of distinctions between clusters. It is stated that the cluster analysis is effective in case when classes of copies are well divided.

Key words: cluster analysis, technique, classification, electronic equipment, electronic component, chip, informative parameter, data centering, dendrogram, dispersive analysis