

УДК 621.382

Г.Ю. Щербакова, В.Н. Крылов

Одеський національний політехнічний університет, Одеса

АВТОМАТИЗАЦІЯ КЛАСТЕРНОГО АНАЛІЗА ІЗМЕРЕНІЙ С ІСПОЛЬЗОВАННЯМ ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЯ

Предложен метод статистического анализа технологических процессов производства изделий радиоэлектронного приборостроения на основе вычисления параметров контрольных карт с помощью помехоустойчивой кластеризации с использованием вейвлет-преобразования. Такой подход позволяет при автоматизированном контроле технологического процесса повысить качество кластеризации в условиях априорной неопределенности, обусловленной отсутствием информации о форме кластеров в случае зашумленных малых наборов данных измерений.

Ключевые слова: контрольные карты, оптимизация, кластеризация, вейвлет-преобразование.

Введение

Постановка проблемы. Важный путь обеспечения качества изделий в радиоэлектронном приборостроении (РП) – управление технологическими процессами (ТП) производства с использованием статистического анализа. ТП считают статистически управляемым при влиянии на него только случайных воздействий. Но при неслучайных воздействиях, например, при износе инструментов, параметры ТП могут существенно отклоняться от целевых значений. Потому одной из основных задач управления ТП является распознавание его перехода в статистически неуправляемое состояние.

Анализ последних достижений и публикаций. Одно из актуальных решений этой проблемы – использование контрольных карт (КК) по количественному признаку [1 – 5]. Эти карты формируют еще на стадии постановки изделия на производство для графического отображения состояния ТП – на них отмечены значения измеряемых параметров изделий во временной последовательности. После определения количества и способа получения выборок, типа КК при обработке результатов измерений выборку делят на однородные подгруппы (кластеры), по параметрам которых определяют значения центральных линий (ЦЛ) и контрольных пределов (КП) КК. На основе такого анализа корректируют ТП, если значения параметров изделий по случайным причинам окажутся вне КП. При разделении выборки на кластеры используется процедура кластеризации [6], позволяющая автоматизировать этот процесс [1, 7, 8]. Одно из преимуществ такого подхода – анализ малых наборов данных измерений, позволяющий понизить стоимость и повысить оперативность отладки ТП, обуславливает существенный недостаток. Он связан с тем, что ошибки оператора, сбои оборудования при измерениях могут обуславливать неопределенность результатов измере-

ний при малом объеме выборок, что эквивалентно их зашумленности. Влияние этих недостатков можно уменьшить, применяя кластеризацию с высоким качеством.

Однако у существующих методов кластеризации в указанных условиях низкое качество [6]. В работе [14] предложен метод кластеризации с использованием вейвлет-преобразования (ВП), позволяющий проводить кластеризацию при высоком уровне помех и малых объемах наборов данных с достаточно высоким качеством.

Цель работы – повысить качество кластеризации с использованием ВП для повышения помехоустойчивости статистического анализа ТП производства с помощью КК по количественному признаку. Для достижения этой цели решены задачи анализа методов кластеризации; разработки процедуры обработки данных при построении КК на примере \bar{X} -R карты; оценки качества кластеризации.

Изложение основного материала

Различают две основные группы методов кластеризации: иерархические и итеративные методы. Общим недостатком иерархических методов является трудоемкость алгоритмов при большом объеме данных анализа, а также, в зависимости от принятой меры расстояния, чувствительность к шуму [6]. В итеративных методах данные перемещаются между кластерами, чтобы минимизировать некоторый функционал качества.

Основные недостатки таких методов – чувствительность к начальной точке поиска и к шуму в данных, отыскание лишь локального минимума. Поскольку обе группы методов чувствительны к шуму данных, и в связи с тем, что кластеризация часто проводится по малым, зашумленным наборам данных и отсутствует априорная информация о форме кластеров, целесообразно при статистическом ана-

лизе ТП применить метод кластеризации с использованием ВП [14], который может позволить повысить качество кластеризации.

Задача кластеризации состоит в разделении данных измерений (образов) на группы (кластеры) с учетом их сходства, которое в метрическом пространстве обычно определяют через расстояние. Расстояние в выбранном подходе к кластеризации рассчитывается от образов к центрам кластеров, координаты которых заранее не известны и находятся одновременно с разделением данных на кластеры. Для снижения влияния указанных недостатков авторами на основе известного метода Я.З. Цыпкина и Г.К. Кельманса [12] разработан метод кластеризации по дисперсионному признаку. Для определения направления движения к экстремуму функционала качества в нем используется метод мультистартовой оптимизации в пространстве ВП (МОВП) [9].

Этот метод реализуется по итеративной схеме

$$c[n] = c[n-1] - \gamma[n] WT_k(Q(x[n], c[n-1])), \quad (1)$$

где $Q(x, c)$ – функционал, который зависит от вектора центров кластеров $c = (c_1, \dots, c_N)$ и от данных измерений $x = (x_1, \dots, x_M)$; $\gamma[n]$ – шаг; n – номер итерации; k – номер старта;

$$WT_k(Q(x[n], c[n-1])) = \{G_{1k}, G_{2k}, \dots, G_{Nk}\} \quad (2)$$

определяет направление движения к экстремуму;

$$G_{jk} = \frac{1}{s_k} \sum_{i=-s_k/2, i \neq 0}^{s_k/2} Q(x[n], c_j + ia) \cdot \Psi_k(i). \quad (3)$$

В (3) s_k – длина носителя вейвлет-функции (ВФ) на k -м старте (s_k – четное число); a – шаг дискретизации; $\Psi_k(i)$ – ВФ на k -м старте (табл. 1); $j = 1, \dots, N$ – размерность вектора параметров.

Шаг $\gamma[n]$ в работе выбирается как $\gamma[1] = 0,4, \dots, 0,6$ [10]. При изменении знака $WT_k(Q(x[n], c[n-1]))$ при переходе через оптимум на $n-1$ шаге $\gamma[n] = 0,5 \cdot \gamma[n-1]$. В результате исследований помехоустойчивости, скорости сходимости и погрешности метода МОВП [9] для оценки направления поиска координаты экстремума в (2) выбраны симметричные и нестационарные ВФ [13]. На первом этапе выбрана ВФ

$$\Psi_1(i) = \begin{cases} 1, & \text{если } i = 1, \dots, \frac{s_1}{2}; \\ -1, & \text{если } i = -1, \dots, -\frac{s_1}{2} \end{cases}$$

и, на следующих этапах, нестационарные ВФ из указанного класса, полученные по схеме (табл. 1), на седьмом этапе направление движения к экстремуму оценивается с помощью

$$\Psi_7(i) = \begin{cases} 1, & \text{если } i = 1; \\ -1, & \text{если } i = -1. \end{cases}$$

Таблица 1

Параметры ВФ для метода МОВП

Название	Обозначение	Значение параметра				
		2	3	4	5	6
Номер старта	k	2	3	4	5	6
Масштаб ВФ	b_k	1	2	3	4	5
Длина носителя	s_k	20	10	6	4	4
Вид ВФ	$\Psi_k(i)$	$\begin{cases} 1/(\alpha_k(i +1)), & \text{если } i > 0, \\ -1/(\alpha_k(i +1)) & \text{если } i < 0, \end{cases} \quad i \in [-s_k/2, +s_k/2], \quad i \neq 0$				

Основные этапы (для двух кластеров ($\gamma = 1, 2$)) приведены ниже.

Этап 1. Формирование обучающей выборки.

Этап 2. Определение количества кластеров на основе известного лр - алгоритма [15].

Этап 3. Оценка функционала качества $Q(x, c)$.

Этап 4. Задаются исходные данные и параметры метода МОВП:

$c[0]$ – начальное приближение к координате оптимума;

$\gamma[1]$ – шаг; вид ВП и ВФ;

a – шаг дискретизации ВФ;

s_1 – длина носителя ВФ первого старта $\Psi_1(i)$;

δ_2 – погрешность кластеризации, которая отвечает требованиям прикладной задачи,

номер старта $k = 1$;

номер итерации $n = 1$.

Этап 5. Оценка направления поиска $WT_k(Q(x[n], c_1[n-1], c_2[n-1]))$ для кластеров по (2) на итерации n .

При этом для каждого из i элементов взвешенной суммы (3) вычисляются характеристические функции $\epsilon_r(x, c_1, c_2)$, $r = 1, 2$:

если $f(x, c_1, c_2) < 0$, то $e_1 = 1, e_2 = 0$,

если $f(x, c_1, c_2) \geq 0$, то $e_1 = 0, e_2 = 1$,

где

$$f(x[n], c_1[n], c_2[n]) =$$

$$\|x[n] - c_1[n]\|^2 - \|x[n] - c_2[n]\|^2.$$

Этап 6. Определение координаты центров кластеров

$$\begin{cases} c_1[n] = c_1[n-1] - \\ \quad - \gamma[n]WT_k(Q(x[n], c_1[n-1], c_2[n-1])); \\ c_2[n] = c_2[n-1] - \\ \quad - \gamma[n]WT_k(Q(x[n], c_1[n-1], c_2[n-1])) \end{cases}$$

при номерах старта $k \leq k_{\max}$, иначе – останов.

Етап 7. Если $\|c_r[n] - c_r[n-1]\| \leq d_1$, поиск на текущем старте заканчивается, иначе – $n = n + 1$ и переход к этапу 5.

Етап 8. Если $k > 1$ и $c_{r,k}^* - c_{r,k-1}^* \leq d_2$ – останов; в противном случае, если $k < k_{\max}$, то увеличивается номер старта $k = k + 1$, выбирается ВФ для оценки направления поиска (см. табл.1) и осуществляется переход к этапу 5.

Тут $c_{r,k}^*$, $c_{r,k-1}^*$ – результаты поиска координат центра r -го кластера на k -м и $k-1$ старте соответственно.

Качество этого метода кластеризации на базе ВП исследовано с использованием известного метода на основе оценки расстояния Хэмминга между тестовой выборкой и исследуемым методом кластеризации при изменении относительной величины среднеквадратического отклонения (СКО) данных в кластерах – q от 0,05 до 0,25; q определялось как

$$q = \frac{q_p}{q_0 \cdot D}.$$

Здесь q_p и q_0 – СКО параметров в кластерах в рабочем режиме и при «обучении», D – расстояние между центрами кластеров обучающей выборки.

В результате исследований по сравнению с известным методом k -средних [6] установлено, что при изменении СКО данных в кластерах качество метода кластеризации с использованием ВП по сравнению с методом k -средних выше в среднем на 0,1.

Контрольные карты для количественных данных позволяют контролировать как расположение центра (уровень, среднее, центр настройки) ТП, так и его разброс (размах, стандартное отклонение) [1 – 5].

Наиболее часто применяют пары карт $\bar{X}-R$ и $\bar{X}-S$ [1]. Значения контролируемого параметра имеют нормальное или близкое к нормальному распределение.

Процедуру обработки данных при построении КК рассмотрим на примере $\bar{X}-R$ карты (\bar{X} – среднее значение, R – выборочный размах для кластера в выборке) [1]. Она используется для управления ТП производства, показатели качества которых представляют собой непрерывные величины (диаметр

сердечника, коэффициент усиления и т.д.). Процедура построения контрольных карт (для одномерного случая) включает следующие этапы.

Етап 1. Формирование обучающей выборки (не менее 100 значений [1]). Построение вариационного ряда значений выборочной характеристики (ВХ). Значения центров кластеров на первой итерации выбираются с учетом минимального и максимального значений членов вариационного ряда. Поскольку разделение на кластеры происходит в одномерном пространстве признаков, кластеризация сводится к последовательному пороговому разделению ВХ на группы.

Етап 2. Определение количества кластеров [15].

Етап 3. Оценка функционала качества $Q(x, c)$.

Етап 4. Задаются исходные данные и параметры метода МОВП согласно этапу 4 метода кластеризации.

Етап 5. Разделяются выборочные характеристики на кластеры.

Етап 6. Вычисляются средние значения \bar{X} для каждого j -го кластера [1]:

$$\bar{X}_j = \sum_{i=1}^{n_j} \frac{x_i}{n_j},$$

где n_j – количество значений выборочной характеристики в кластере; j – номер кластера; x_i – значение выборочной характеристики.

Етап 7. Вычисляется общее среднее значение $\bar{\bar{X}}$ (центральная линия \bar{X} КК) [1]

$$\bar{\bar{X}} = \sum_{j=1}^M \frac{\bar{X}_j}{M},$$

где M – количество кластеров.

Етап 8. Вычисляется размах для каждого кластера [1]

$$R_j = x_{j\max} - x_{j\min},$$

где $x_{j\max}$, $x_{j\min}$ – соответственно максимальное и минимальное значение выборочной характеристики в кластере.

Етап 9. Вычисляется средний размах (центральная линия R КК) [1]

$$\bar{R} = \sum_{j=1}^M \frac{R_j}{M}.$$

Етап 10. Вычисление значений контрольных пределов проводятся по методике [1, 7]:

для \bar{X} – карты – как $\bar{\bar{X}} \pm A_2R$;

для R – карты – как D_3R , D_4R .

Значения зависящих от объема выборки A_2 , D_3 , D_4 приведены в [4].

Выводы

Проведен анализ методов кластеризации, применяемых при статистическом анализе ТП производства ИЭТ, с учетом их помехоустойчивости. Разработаны и обоснованы процедуры обработки данных при построении КК на примере \bar{X} -R карты в процессе статистического анализа ТП с применением метода кластеризации с использованием ВП.

Проведены экспериментальные исследования для оценки качества кластеризации по сравнению с известным методом k-средних и установлено, что при изменении СКО данных в кластерах качество метода кластеризации с использованием ВП по сравнению с методом k-средних выше в среднем на 0,1.

Эти результаты позволяют рекомендовать разработанную процедуру построения КК в процессе статистического анализа ТП в радиоэлектронном приборостроении при неопределенности результатов измерений, обусловленных малым объемом выборок.

Список литературы

1. Болтян А.В. Прогнозирование и оценка параметров продукции / А.В. Болтян, И.А. Горобец – Донецк: Норд-Пресс, 2004. – 192 с.
2. Статистические методы. Статистическое управление качеством. Термины и определения (ISO 3534.2-93):ГОСТ Р50779.11 – 2000. – [Начало действия 2001-07-01]. М.: Госстандарт России 2000.
3. Статистические методы. Контрольные карты. Общее руководство и введение. (ISO 7870-93) ГОСТ Р50779.40-96. – [Начало действия 1996-07-01]. М.: Госстандарт России 1997.
4. Статистические методы. Контрольные карты Шухарта. (ISO 8258-91) ГОСТ Р50779.42-99. – [Начало действия 1999-04-13]. М.: Госстандарт России 1999. – 36 с.
5. Статистические методы. Индексы возможностей процессов. ГОСТ Р50779.44-2001. – [Начало действия 2002-07-01]. М.: Госстандарт России 2001.

6. Дуда Р. Распознавание образов и анализ сцен / Р. Дуда, П. Харт. – М.: Мир, 1976. – 509 с.

7. Жулинский С.Ф. Статистические методы в современном менеджменте качества / С.Ф. Жулинский, Е.С. Новиков, В.Я. Поспелов – М.: Фонд «Новое тысячелетие», 2001. – 208 с.

8. Контроль качества с помощью персональных компьютеров / Т. Макино, М. Охаси, Х. Докэ, К. Макино; Пер. с яп. А.Б.Орфенова. Под ред. Ю.П.Адлера. – М.: Машиностроение, 1991. – 224 с.

9. Крилов В.Н. Субградієнтний ітеративний метод оптимізації в просторі вейвлет-перетворення./ В.Н. Крилов, Г.Ю. Щербакова // Збірн. наук. праць Військового інституту Київського нац. ун-ту ім. Т. Шевченка. – К., 2008. – Вип. 12. – С. 56 – 60.

10. Полак Э. Численные методы оптимизации / Э. Полак. – М.: Мир, 1976. – 509 с.

11. Статистические методы повышения качества: Пер. с англ. и доп. Ю.П.Адлера, Л.А.Коняревой / Под ред. Х. Кумэ. – М.: Финансы и статистика, 1990. – 304 с.

12. Цыпкин Я.З. Адаптация и обучение в автоматических системах. / Я.З. Цыпкин – М.: Наука, 1968. – 400 с.

13. Krylov V.N. Contour images segmentation in space of wavelet transform with the use of lifting / V.N. Krylov, M.V. Polyakova // Optical-electronic informatively-power technologies. – 2007. – № 2 (12). – P. 48 – 58.

14. Shcherbakova G. Electronic Apparatus Automation Inspection with Adaptive Clustering in Wavelet Domain. / G. Shcherbakova, V. Krylov, S. Antoshchuk // Досвід розробки та застосування приладо-технологічних САПР в мікроелектроніці: міжнар. наук.-техн. конф., 24–28 лют. 2009 р., тези доп. – Львів, 2009. – С. 153 – 154.

15. Крисилов В.А. Решение задачи таксономии на основе гипотезы компактности при анализе данных / В.А. Крисилов, Н.О. Блудян, С.А. Юдин // Искусственный интеллект. – 2005. – № 4. – С. 699 – 707.

Поступила в редколлегию 16.12.2014

Рецензент: д-р техн. наук, проф. С. Г. Антошук, Одесский национальный политехнический университет, Одесса.

АВТОМАТИЗАЦІЯ КЛАСТЕРНОГО АНАЛІЗУ ВИМІРЮВАНЬ З ВИКОРИСТАННЯМ ВЕЙВЛЕТ-ПЕРЕТВОРЕННЯ

Г.Ю. Щербакова, В.Н. Крылов

Запропонований метод статистичного аналізу технологічних процесів виробництва виробів радіоелектронного приладобудування на основі обчислення параметрів контрольних карт шляхом завадостійкої кластеризації з використанням вейвлет-перетворення. Такий підхід дозволяє при автоматизованому контролі технологічного процесу підвищити якість кластеризації в умовах апріорної невизначеності, яка обумовлена відсутністю інформації про форму кластерів в умовах зашумлених малих наборів даних вимірювань.

Ключові слова: контрольні карти, оптимізація, кластеризація, вейвлет-перетворення.

AUTOMATION OF THE CLUSTER ANALYSIS OF MEASURING WITH USE OF WAVELET TRANSFORMING

G.Y. Shcherbakova, V.N. Krylov

The method of statistical technological process analysis for electronic apparatus production is suggested. This method based on noise stabilities clustering with wavelet transforming utilization. Such approach allows inspected of technological production process of electronic apparatus with highest level clustering quality. Such approach is recommended in case of control data with high level noise, when the information about form of clusters is absent.

Keywords: checklists, optimization, clustering, wavelet transform.