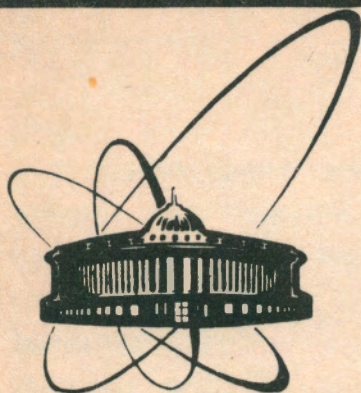


92-143



сообщения
объединенного
института
ядерных
исследований
дубна

P10-92-143

Ю.Н.Пепельшев, В.Дзвинель

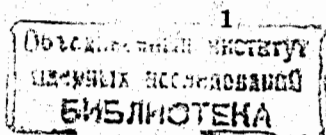
ВОЗМОЖНОСТЬ ОПЕРАТИВНОГО КОНТРОЛЯ
КАЧЕСТВА ПРОДУКЦИИ С ИСПОЛЬЗОВАНИЕМ
МЕТОДОВ КЛАССИФИКАЦИИ
МНОГОПАРАМЕТРИЧЕСКИХ ДАННЫХ

1992

1. Введение

Во многих сферах производства для обеспечения требуемого качества (Q) продукции необходимо контролировать величину некоторого числа (n) технологических параметров X_i ($i=\overline{1, n}$) в пределах допустимых отклонений от их средних \bar{X}_i (оптимальных) значений. Различные вариации изменения величины контролируемых параметров даже в пределах допустимых отклонений, естественно, создадут некоторую вариацию и в качестве выпускаемой продукции. Если эти вариации можно оценить в процессе производства, то уже на различных стадиях изготовления продукции можно оценить её качество. Как это можно сделать? Допустим, что за время производства продукции (T) контроль параметров осуществляется в моменты времени $t=j\Delta t$, $j = \overline{1, k}$, $T = k\Delta t$. Таким образом, текущий j -й цикл контроля характеризуется набором данных $X_{nj} = (x_1, x_2, x_3, \dots, x_1, \dots, x_n)$, а полный набор за время T - множеством $\{X_{nk}\}$. В общем случае качество есть некоторая функция $Q(X_{nj})$, которую можно описать теоретически или найти с помощью эмпирических моделей. Такой подход к контролю качества будет ограничен в следующих случаях:

- При невозможности получить аналитическое выражение или построить регрессионную модель для $Q(X_{nj})$.
- Бессмысленность получения аналитического выражения в случае потери точности с ростом числа параметров.
- При отсутствии данных о влиянии каких-либо параметров на Q или данных о их корреляции с другими параметрами.



Практически при достижении некоторого числа параметров контроль качества, основанный на аналитическом представлении функции $Q(X_{nj})$, становится непригодным. Кроме того, если характер изменения параметров имеет статистическую природу, что является наиболее общим случаем, то для оценки Q необходимо использовать также и статистические методы анализа данных. Использование традиционных методов математической статистики в случае многомерной задачи не всегда даёт надёжные и удобные в интерпретации результаты. Существуют и другие ограничения на применимость обычных методов математической статистики к анализу X_{nj} .

■ Трудность многомерного анализа в случае, если статистические характеристики измеряемых параметров различаются по величине в несколько раз.

■ Проблема оценки Q в реальном масштабе времени.

В работе предлагается иной подход к анализу параметров X_{nj} . В нём существенно устраняются ограничения, связанные с использованием традиционных аналитико-статистических методов анализа данных.

2. Концепция анализа многомерных данных в задаче

контроля качества.

Основу предлагаемого метода анализа многомерных данных составляет сочетание методов распознавания образов с визуализацией многомерных структур. Измеренные параметры X_{nj} можно представить в виде вектора (точки) в n -мерном евклидовом пространстве: $\vec{X}_{nj} = (x_1, x_2, \dots, x_1, \dots, x_n)_j^T$ (1) где значения параметров $X_{nj}, i=1, n$ есть координаты вектора \vec{X}_{nj} в j -й момент времени (здесь T — знак транспонирования). Вектор \vec{X}_{nj} рассматривается как текущий образ технологического

процесса. Текущий образ добавляется к ранее зарегистрированным, совокупность которых образует базу данных. На каждом j -м этапе измерения совокупность образов анализируется с помощью методов распознавания с целью выделения характерных кластерных структур. В данном случае кластер (класс) есть совокупность образов, объединенных некоторыми общими свойствами. Исследование классов позволяет найти образы, отображающие нормальные условия протекания технологического процесса, и образы, характерные для различных отклонений от этого режима. В решении этой задачи может быть и иной подход: зная классы нормального и аномального режимов (классы разного типа) можно определить, к какому классу принадлежит текущий образ. Появление отклонений в структуре кластеров от структуры, соответствующей нормальному технологическому режиму, характеризуется изменением формы, смещением центра текущего кластера или появлением нового кластера. Последнее свидетельствует о существенных изменениях, происходящих в технологическом процессе. Причину изменения кластерной структуры можно установить опытным путём или с помощью взаимно-корреляционного анализа между измеряемыми параметрами. В любом случае, если контролируемый процесс обеспечивает представление исходных данных в виде (1) и последние группируются по кластерам, то указанное свойство процесса можно использовать для контроля за ходом самого процесса.

2.1 Способы формирования вектора измеряемых параметров

В зависимости от времени производства (T) некоторого вида

продукции и числа контролируемых параметров (n) можно использовать несколько вариантов контроля:

1. По времени.

$$\vec{X}_i, \quad X_i = (x_{i1}, x_{i2}, \dots, x_{ij}, \dots, x_{ik})^T, \quad j = \overline{1, k}, \quad i - \text{номер параметра.}$$

2. По параметрам.

$$\vec{X}_j, \quad X_j = (x_{1j}, x_{2j}, \dots, x_{ij}, \dots, x_{nj})^T, \quad i = \overline{1, n}, \quad j - \text{й момент времени.}$$

3. Комбинированный (по времени и по параметрам)

$$\vec{X}_{ij}, \quad j = \overline{1, k}, \quad i = \overline{1, n}.$$

$$X_{ij} = \begin{vmatrix} x_{11} & x_{12} & x_{13} & \dots & x_{1j} & \dots & x_{1k} \\ x_{21} & x_{22} & x_{23} & \dots & x_{2j} & \dots & x_{2k} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & x_{n3} & \dots & x_{nj} & \dots & x_{nk} \end{vmatrix}$$

4. Периодическое повторение комбинированного варианта.

$$\vec{X}_{ij}^m, \quad j = \overline{1, k}, \quad i = \overline{1, n}, \quad t = \Delta t * j, \quad T_c = \Delta t * k, \quad m = \overline{1, L}, \quad L = \frac{T}{T_c},$$

где T_c - время цикла измерения.

2.2. Формулировка и способы реализации задачи распознавания

образов с помощью признаков.

По сути решается задача классификации данных по некоторым в общем случае заранее неизвестным инвариантным признакам. Исходное множество образов состоит из неизвестного числа (M) подмножеств (классов), которые представляют собой выборки из разных генеральных совокупностей. Задача классификации состоит в выделении классов $\omega_j, j = \overline{1, M}$, характеристики

которых можно было бы адекватно связать с основными состояниями технологического процесса: $Q_j \leftrightarrow \omega_j$. В такой формулировке исходная задача не имеет эффективных методов решения. Ограничения в основном вызваны неизвестным числом классов и неизвестной степенью их пересечения (P). Для удобной интерпретации физических результатов в условиях априорной неопределённости в значениях M и P достаточно выделить основные взаимно-компактные классы, обладающие слабым пересечением и образующие выпуклые области в исходном пространстве признаков (координат). Сравнительный анализ наиболее приемлемых для решения этой задачи алгоритмов показал, что удачное сочетание формализованных и неформализованных процедур анализа имеет неиерархический алгоритм ISODATA [1, 2, 3] (итеративный самоорганизующийся метод анализа) и иерархический алгоритм MNN [4]. В соответствии с алгоритмом ISODATA разделение образов по классам осуществляется путём минимизации критерия взаимной компактности классов $K(\omega_1, \omega_2, \dots, \omega_M)$, (M - число классов), в качестве которого использовалась сумма квадратов ошибки:

$$K(\omega_1, \omega_2, \dots, \omega_M) = \sum_{j=1}^M d_j^2, \quad d_j^2 = \frac{1}{N_j} \sum_{i=1}^n \sum_{\alpha=1}^{N_j} (X_{ik} - Z_{jk})^2 - \text{средний}$$

квадрат евклидоваго расстояния между образами класса ω_j и его центром \vec{Z}_j , N_j - число образов в классе ω_j , n - число координат образа. Здесь j, i, k - текущие индексы. Отметим некоторые особенности программной реализации и конкретного применения алгоритма ISODATA. Входные параметры алгоритма, θ_s и θ_c , задающие условия разъединения j -го класса $\sigma_{j\max} > \theta_s$, где $\sigma_{j\max}^2$ - максимальная дисперсия образов относительно центра j -го класса, и объединения j -го и i -го классов $[(\vec{Z}_j - \vec{Z}_i)^T (\vec{Z}_j - \vec{Z}_i)]^{1/2} < \theta_c$ варьируются в расчёте до получения

минимума $K(\omega_1, \omega_2, \dots, \omega_M)$ при заданном ожидаемом числе классов. По методу MNN для векторов \vec{X}_i в евклидовом пространстве R $\dim R = n$, $\vec{X}_i = (x_{i1}, x_{i2}, \dots, x_{in})^T$, $i = \overline{1, N}$, N - число векторов, определяются два типа расстояния. Первое: $S_{ij} = (X_i - X_j)^T (X_i - X_j)$ - классическое евклидово расстояние. Второе: $mnv_{ij} = L_1(j) + L_j(i)$, где $L_1(j)$ означает положение точки X_j в списке k_1 соседей точки X_i и выбирается в соответствии с увеличением расстояния S_{im} , где m -номер соседней точки, определяемый из условия $S_{im} \leq R_{cut}$ (2). Если число ближайших соседей k_1 , удовлетворяющих условию (2), различно для каждой i -ой точки, то максимальное mnv расстояние меньше или равно

$mnv_{max} = \max_i \{k_i\} + \max_i \{ \{k_i\} - \{ \max \{k_i\} \} \}$, и можно заключить: $\forall i, j; 2 \leq mnv_{ij} \leq mnv_{max}$. Таким образом, в методе MNN определяющим фактором классификации является радиус отсечки R_{cut} .

2.3. Необходимость визуализации данных и техника преобразования многомерных структур

Нужно отметить, что численные характеристики кластеров, с помощью которых можно оценить изменение сложных гиперструктур, не всегда дают наглядную информацию о ходе контролируемого процесса. Поэтому в данном случае оказалось целесообразным исходную совокупность образов "свёрнуть" из n -мерного в 2 - или 3 - мерное пространство для её визуального контроля. Процедура визуализации, дополняющая предыдущий анализ, обогащает информацию о n -мерной структуре кластеров и позволяет более оперативно и более надёжно осуществлять контроль. Используются несколько методов преобразования многомерных структур в двух - или трёх -

мерное пространство. Первый основан на преобразовании Карунена - Лоэва [5], второй представляет собой дальнейшее развитие известной концепции Неймана [6], третий базируется на "Simulated annealing"-методологии [7]. Если R есть евклидово пространство размерностью n $\dim R = n$, $n < n$ и точка в n - мерном пространстве представлена в виде $X_i = (x_{i1}, x_{i2}, \dots, x_{in})$, где $i = \overline{1, N}$, N - число точек, то для сохранения подобия структуры кластеров $R_0 = \{x_i; i = \overline{1, N}\}$ (особенно расстояния между точками), при переходе в пространство R меньшей размерности должно, согласно [7], выполняться условие

$$E(R, R_0) = \min_R E(R, R), \text{ где}$$

$$E(R, R_0) = \sum_{k < j} (x_{jk} - x'_{jk})^2 / \sum_{k < j} x_{jk}, \quad \sum = \sum_{j=2}^N \sum_{k=1}^{j-1}$$

Ниже представлено краткое описание системы контроля (диагностирования) технологических процессов, созданной на основе вышеописанной концепции.

3. Общее описание системы контроля технологических процессов

Детальное описание системы контроля представлено в работе [8]. Система построена на базе ЭВМ РС/АТ с входными измерительными блоками типа PCLAB-712 (аналого-цифровой преобразователь на 16 входов). Принципиальная блок-схема аппаратной части системы показана на рис.1. Существующие возможности блоков LABCARD позволяют регистрировать сигналы с широкого класса датчиков (например, датчиков температуры, давления, расхода, вибраций, перемещений и т.д.). С точки зрения используемой в системе процедуры анализа данных,

физическая природа измеряемых процессов не является значимым фактором. Поэтому в систему могут быть введены любые параметры, важные для контроля за ходом некоторого технологического процесса. Предусмотрено два режима работы: основной и расширенный. Первый режим используется для автоматического контроля, второй – для проведения детальных исследований измеряемых параметров.

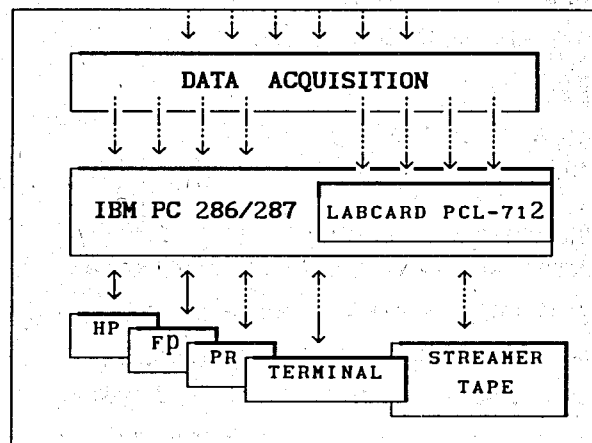


Рис. 1. Принципиальная блок-схема системы контроля технологических процессов

ХАРАКТЕРИСТИКИ СИСТЕМЫ

- | | |
|---|----------|
| ■ число входных сигналов | 8(16) |
| ■ диапазон изменения входных сигналов | +/-5(1)V |
| ■ максимальное число отсчётов для каждого сигнала | 8192 |
| ■ максимальная частота измерения | 50KHz |
| ■ максимальная размерность конечного массива данных | 200*256 |

3.1. Программное построение системы

Программная часть системы построена в виде пакета программ с общей организующей программой, работающей по типу разветвлённого меню. Программы написаны на языках FORTRAN 77 и PASCAL. Ниже даны функции основных команд меню.

3.2. Накопление данных

Выбранные n сигналов "одновременно" регистрируются с заданной частотой с накоплением k значений по каждому из измеряемых входных параметров. В конце измерения данные записываются на твёрдый диск для последующей обработки.

3.3. Предварительная обработка данных

Выполняется при исследовании статистических характеристик контролируемых процессов. Вычисляются спектральная плотность мощности, функции когерентности и фазовый сдвиг между i - м и j - м процессами, а также плотность распределения вероятности значений параметров. Вычисляются также и другие статистические характеристики, образующие стандартный набор результатов статистического анализа временных рядов.

3.4. Кластерный анализ

Для анализа кластерной структуры совокупности векторов \vec{x}_{nj} используются два дополняющих друг друга алгоритма распознавания образов: ISODATA и MNN. Вычисляются также энтропия, дисперсия элементов, входящих в кластер, и другие характеристики, необходимые для описания выделенных кластеров и определения тенденции их изменения.

3.5. Визуализация

Оператор может наблюдать на экране дисплея распределение образов по кластерам, статистику каждого кластера, информацию о текущем векторе \vec{x}_{nj} и другие данные.

4. Пример использования системы для контроля за состоянием ядерного реактора

Результаты работы системы можно проиллюстрировать на примере её использования в качестве системы контроля общего состояния импульсного реактора периодического действия ИБР-2 (Дубна). Одной из характеристик качества работы реактора является стабильность формы спектра флуктуаций мощности. Поэтому образом реактора, характеризующим его состояние, была выбрана спектральная плотность колебаний мощности:

$$S_w = |F[W(\Delta t * i)]|^2 = (S_{w1}, S_{w2}, \dots, S_{wn}), i = \overline{1, k};$$
 n - число значений спектральной плотности, F - Фурье-преобразование временной зависимости изменения мощности, зарегистрированной с интервалом Δt в k точках. При изменении установившихся флуктуаций внешней реактивности (флуктуаций источников возмущений мощности) форма спектра колебаний мощности соответствующим образом изменяется. Это отражается на кластерной структуре спектров. В данном случае размерность исходного пространства, в котором анализируется вектор \vec{S}_w , определялась числом значений спектральной плотности (n) и была равна 256. На рис.2 показано последовательное изменение формы спектров колебаний мощности в период работы реактора с 1989 по 1991 г. Несмотря на кажущееся хаотическое изменение формы спектров, при их кластеризации выделились две характерные устойчивые структуры, центры которых показаны на рис.3.

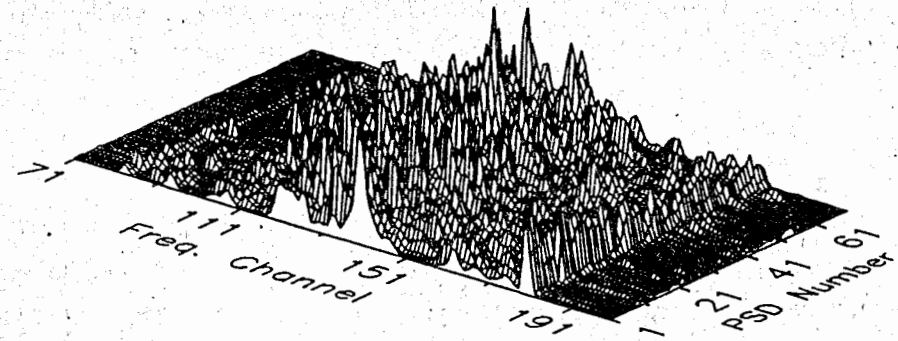


Рис.2.Изменение формы спектра колебаний мощности реактора ИБР-2. По оси частот отложен номер частотного интервала, по другой оси - номер спектра в последовательном ряде измерений.

Соответствующее распределение образов по кластерам, "свёрнутое" из 256 - мерного пространства в 2 - мерное, приведено на рис.4. Как видно из рис.3, весьма незначительные, но устойчивые (здесь сохраняющиеся при случайных изменениях) различия в форме спектров (в среднем $\leq 1\%$) приводят к чёткому разделению спектров даже в 2 - мерном представлении. Оба шумовые состояния реактора соответствуют нормальному эксплуатационному режиму, но различаются незначительными отклонениями в уровне вибраций одной из технологических систем реактора. Изменение

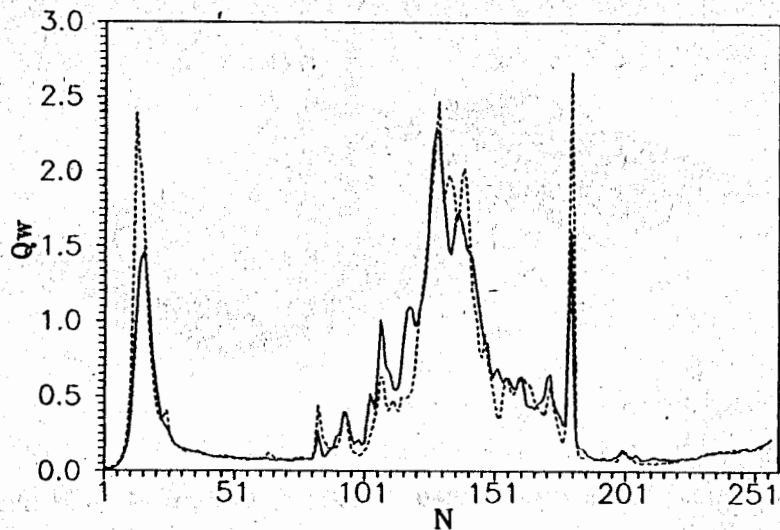


Рис. 3. Форма спектров колебаний мощности реактора ИБР-2, соответствующая центрам двух кластеров. По оси абсцисс отложен номер частотного интервала.

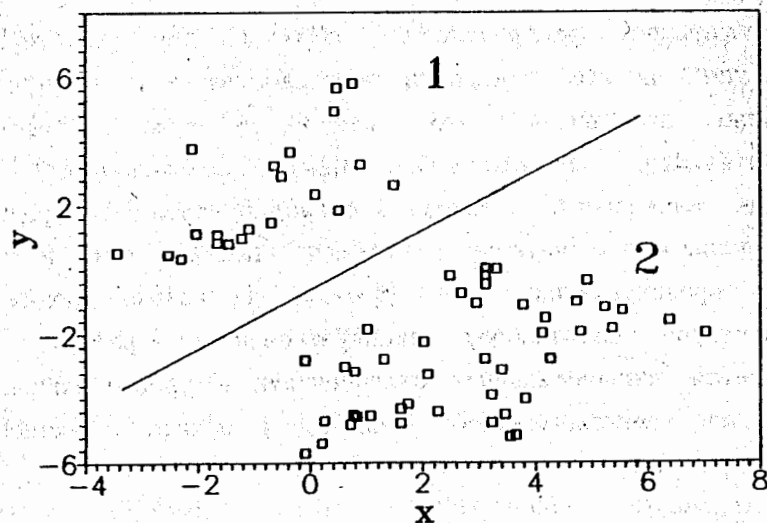


Рис. 4. Распределение спектров колебаний мощности по кластерам, свёрнутое из 256 - мерного в 2 - мерное пространство.

относительных флуктуаций мощности ($\Delta W/W$), вызывающее переход реактора из одного шумового состояния в другое (из одного кластера в другой), оценивается на уровне $\approx 10^{-2}$.

5. Заключение

Идея контроля качества некоторой продукции на основе вышеописанной концепции возникла у авторов в процессе использования созданной ими системы диагностирования импульсного ядерного реактора ИБР-2 (Дубна). Поскольку в основу используемых процедур анализа данных входят лишь формализованные процедуры, то физическая природа контролируемого объекта в данном случае не имеет принципиального значения, хотя и может учитываться в процессе кластеризации. Существенным фактором успешного применения такой системы для контроля качества продукции являются в основном лишь свойства контролируемых процессов, особенно их способность представляться в виде некоторой кластерной структуры. Обычно, если измеряемые параметры выбраны "правильно", т.е. они действительно влияют на состояние контролируемого объекта (в данном случае качество продукции), то кластеризация данных будет заметна в той или иной форме. И поскольку метод очень чувствителен даже к небольшим и скрытым в шумах изменениям параметров ($\Delta X/X \approx 10^{-3}$), то, как показывает практика, во многих случаях этого оказывается вполне достаточно для детальной градации продукции по качеству.

Литература

- 1 M.R.Anderberg, *Cluster Analysis for Applications*, Academic Press, NY, 1973.
- 2 R.C.Gonzalez D.N.Fry, R.C. Kryter, Results in the Application of Pattern Recognition Methods to Nuclear Reactor Core Component Surveillance, *IEEE Trans. Nucl.Sci.*, 21 (1974) 750.
- 3 M.Bargel, J.Kitowski, Ju.N.Pepolyshv, Commun. JINR Dubna, 13-88-345, 1988.
- 4 K.Chidananda, G.Krishna, Agglomerative Clustering Using the Concept of Mutual Nearest Neighborhood, *Pattern Recognition*, 102 (1978) 105.
- 5 J.T.Tou, R.C.Gonzalez, *Pattern Recognition Principles*, Addison-Westly Publishing Company, 1974, p.290.
- 6 H.Niemann, Linear and Nonlinear Mapping of Patterns, *Pattern Recognition*, 122 (1980) 83.
- 7 S.Kirkpatrick, C.D.Jr. Gellatt, M.P.Vecchi, *Science*, 220 (1983) 671.
- 8 Ju.N.Pepolyshv, W.Dzwinel, Pattern Recognition System for the Nuclear Reactor Noise Image Analysis and Diagnostics, in: Proceedings of the International Symposium on Nuclear Reactor Surveillance and Diagnostics, SMORN-VI, Gatlinburg, Tennessee, USA, May 19-24, 1991, V.2, pp.61.01-61.17.

Рукопись поступила в издательский отдел
1 апреля 1992 года.