УДК 658.562:621 (035)

Щапов П.Ф.

ПРИМЕНЕНИЕ ТЕСТОВЫХ СТАТИСТИК В СИСТЕМЕ АВТОМАТИЧЕСКОГО ИЗМЕРИТЕЛЬНОГО КОНТРОЛЯ ТЕХНОЛОГИЧЕСКИХ ПРОЦЕССОВ

Обнаружение изменений свойств случайных сигналов и динамических систем является важной прикладной задачей технической и медицинской диагностики, контроля стабильности и точности технологических процессов, анализа функциональной нестационарности временных рядов [1]. Прикладная статистика, используемая для решения таких задач, в подавляющем большинстве случаев базируется на параметрическиом тестировании числовых характеристик случайных информационных процессов, игнорируя функциональные неслучайные модели изменения этих характеристик. Такое игнорирование приводит к потере потенциальной информации и снижает достоверность выводов о наличии или отсутствии изменений в контролируемой динамической системе. Более того, такая невостребованная информация могла бы помочь в количественной оценке неслучайных параметрических изменений и в классификации причин, вызвавших то или иное технологическое нарушение.

Особую важность вопросы контроля таких нарушений приобретают для энергоемких производств, для которых любые технологические отклонения сопровождаются повышенным энергопотреблением.

Классификационные многофакторные методы статистического анализа случайных последовательностей и временных рядов используют линейные модели классификации [2,3], для которых уменьшение остаточной дисперсии обусловлено расширением модели. Такое расширение позволяет вводить дополнительные параметры в исходную линейную модель, увеличивая число информативных слагаемых дисперсионного разложения, поскольку остаточная энтропия результатов наблюдения уменьшается (из-за снижения остаточной дисперсии).

К сожалению, проверка альтернативных гипотез в рамках дисперсионного синтеза параметрических моделей затруднена и использование таких моделей возможно с учетом ошибки только первого рода и лишь для проверки основной (простой или сложной) гипотезы.

Более привлекательным для классификации состояний объекта контроля являются модели дискриминантного анализа[4], однако нарушение априорных предположений в виде такой модели резко снижает достоверность принятия решений, хотя в этом случае для оптимизации модели используют вероятности ошибок как первого, так и второго рода.

Целью данной работы является раскрытие возможностей дисперсионного анализа регрессионных моделей, когда для получения информации о состоянии объекта контроля используют тестовые статистики односторонней классификации при двумерных наблюдениях. Такие статистики позволяют проводить эффективный контроль режимов технологических линий по производству растительных масел.

Вероятностная модель информационного сигнала. Достаточно общей математической (статистической или вероятностной) моделью служит представление сигнала уравнением [3]:

$$Y_t = f(t) + U_t \cdot t = 1,2,...N\,, \tag{1}$$

где *f(t)* – детерминированная (систематическая) последовательность наблюдений; U_t – случайная последовательность; *t* – момент наблюдения; *N* – общее число равноотстоящих друг от друга на оси времени моментов, наблюдений.

Условимся, что влияние времени *t* сказывается на изменении только систематической составляющей. Таким образом временной ряд (1) является аналогом непрерывного, нестационарного по математическому ожиданию, случайного процесса с дискретным временем наблюдения. Остаточная дисперсия σ^2 последовательности U_t постоянна во времени. Будем рассматривать *f(t)* как возрастающую или убывающую функцию времени (тренд), так как предметом исследования являются переходные процессы изменения информационных сигналов технологического контроля, вызванные планируемым скачкообразным изменением режима производства, например, запуском технологической линии.

Дисперсионный анализ вероятностной модели. Разобьем общее время наблюдения на K последовательных интервалов с n_j числом отсчетов ($j = \overline{1,K}$) в каждом таком интервале. Тогда общее число отсчетов равно:

$$N = \sum_{j=1}^{K} n_j\,. \tag{2}$$

Используем кусочно-линейную аппроксимацию функции *f(t)*, представляя ее на *K* интервалах наблюдения линейными регрессиями:

$$E[Y_j / t] = A_j + B_j \cdot t\,, \tag{3}$$

где A_j, B_j – оценки параметров α_j, β_j модели наблюдения сигнала Y_t:

$$Y_{ji} = \alpha_j + \beta_j \cdot t_{ji} + Z_{ji,}, (j = \overline{1,K}, i = \overline{1,n_j})\,, \tag{4}$$

где z_{ji} – случайный остаток с дисперсией σ^2.

Рассмотрим основную линейную гипотезу порядка 2(*K*-1).

$$\text{Но: } (\alpha_1 = \alpha_2 = ... = \alpha_k; \beta_1 = \beta_2 = ... = \beta_k). \tag{5}$$

Согласно [4] её можно разбить на три независимых части:

$$H_0^{(1)} : \beta_1 = \beta_2 = ... = \beta_k;$$

$H_0^{(2)}$: групповые средние лежат на прямой;

$H_0^{(3)}$: угловой коэффициент этой прямой равен β_c, причем $\beta_{\tilde{n}} = \beta_1 = ... = \beta_k$.

Для проверки выдвинутых гипотез используем разложение суммы *S* квадратов отклонений наблюдений Y_{ji} от общего среднего $\overline{Y}$ на пять слагаемых [4]:

$$S = S_0 + S_{WG} + S_G + S_W + S_R\,, \tag{6}$$

где: $S_0 = w_0 B_0^2$; $S_{WG} = \dfrac{w_c w_m}{w_0}(B_c - B_m)^2$; $S_G = \sum_j n_j [\bar{Y}_j - \bar{Y} - B_m(\bar{t}_j - \bar{t})]^2$;

$S_W = \sum_j w_j (B_j - B_c)^2$; $S_R = \sum_j \sum_i [\bar{Y}_{ji} - \bar{Y}_j - B_j(t_{ji} - \bar{t})]^2$.

Постоянные w_0, w_m и w_c связаны уравнением:

$$w_0 = w_m + w_c, \text{ где}$$

$$w_0 = \sum_j \sum_i (t_{ji} - \bar{t})^2,$$

$$w_m = \sum_j n_j (\bar{t}_j - \bar{t})^2,$$

$$w_c = \sum_j \sum_i (t_{ji} - \bar{t}_j)^2,$$

t_{ji} – время отсчета значения Y_{ji}.

Угловые коэффициенты B_0, B_m и B_c определяются выражениями:

$$B_c = E[\,B_j\,];$$

B_0 – угловой коэффициент регрессии, построенной по всему общему временному ряду;

$$B_m = w_m^{-1}(w_0 B_0 - w_c B_c).$$

Тестовые статистики для проверки гипотез $H_0, H_0^{(1)}, H_0^{(2)}$ и $H_0^{(3)}$ имеют вид:

$$\begin{cases} F_{0\Sigma} = (S_{WG} + S_G + S_W)/[\bar{S}_R \cdot 2(K-1)], & \text{гипотеза } H_0; \\ F_0^{(1)} = S_W/[\bar{S}_R \cdot (K-1)], & \text{гипотеза } H_0^{(1)}; \\ F_0^{(2)} = S_G/[\bar{S}_R \cdot (K-2)], & \text{гипотеза } H_0^{(2)}; \\ F_0^{(3)} = S_{WG}/\bar{S}_R, & \text{гипотеза } H_0^{(3)}; \end{cases} \tag{7}$$

где $\bar{S}_R = S_R/(N-2K)$ – оценка остаточной дисперсии σ^2.

Формирование пространства информационных признаков. Представим, формально, задачу обнаружения технологических нарушений, как задачу классификации, решаемую в рамках дискриминантного анализа [4].

Введем два класса π_0 и π_1, характеризующих соответственно и ненормальное (с нарушениями) состояние объекта контроля принятия решений:

$$\gamma_0, \text{ если } \pi \in \pi_0,$$
$$\gamma_1, \text{ если } \pi \in \pi_1.$$

В качестве информационных признаков $X_1,\ldots,X_q$ используем F-статистики (7), учитывая, что независимыми являются лишь статистики $F_0^{(1)}$, $F_0^{(2)}$, $F_0^{(3)}$ ($q \le 3$, случай одного информационного сигнала). Если число информационных сигналов l > 1, то $q \le 3l$.

При выборе информационных признаков следует выбирать признаки с максимальной дискриминирующей способностью (имеющих максимальную вероятность правильной классификации). Кроме этого, следует учитывать, что *F*-статистики (7) являются случайными величинами с центральным (если $\pi \in \pi_0$) или нецентральным (если $\pi \in \pi_1$) *F*-распределением [4]. Более того, как для класса π_0, так и для класса π_1 *F*-статистики могут иметь нецентральное *F*-распределение с различными параметрами нецентральности λ_s. Это означает, что переход от одного класса состояний к другому сопровождается изменением как средних значений признаков $X_1,\ldots,X_q$ так и их дисперсий. Последнее означает, что вероятность ошибок первого (α) и второго (β) рода неодинаковы. Пусть признак X_s, $s=\overline{1,q}$, характеризуется по классам π_0 и π_1 средними значениями $\mu_s^{(0)}$ и $\mu_s^{(1)}$ и дисперсиями $D_s^{(0)}$ и $D_s^{(1)}$. В этом случае, для условных по классам π_0 и π_1 нецентральных F-распределений соотношение между вероятностями ошибок α и β определяются неравенствами:

$$\alpha >> \beta,\ \text{если}\ \ \mu_s^{(0)} < \mu_s^{(1)}, \tag{8}$$

$$\alpha << \beta,\ \text{если}\ \ \mu_s^{(0)} > \mu_s^{(1)} \tag{9}$$

Условия (8) и (9) указывают на необходимость использования для классификации четного числа признаков, половина из которых удовлетворяет, например, условию (8), а вторая половина – условию (9). В этом случае обеспечивается условие $\alpha \approx \beta$, что соответствует увеличению мощности правила классификации.

Выбор дискриминантной функции. Для выбора решений γ_0 или γ_1 удобнее всего воспользоваться отношением максимального правдоподобия:

$$\Omega = \frac{\prod_{s=1}^{q} f(X_s/\pi_0)}{\prod_{s=1} f(X_s/\pi_1)}, \tag{10}$$

где $f(X_s/\pi_r)$ – условная плотность распределения вероятностей признака X_s при $\pi \in \pi_r$, $r=\overline{0,1}$.

Если $f(X_s/\pi_r)$ является законом Гаусса, то есть $X_s \sim N(\mu_s^{(r)}, D_s^{(0)})$, то отношение (10) превращается в квадратичную дискриминантную функцию [4]. Исследование влияния нарушения априорных предположений о нормальности плотностей $f(X_s/\pi_r)$, $s=\overline{1,q}$, показало, что это влияние тем меньше:

а) чем сильнее пересекаются классы π_0 и π_1;

б) чем больше число информационных признаков;

в) чем больше параметры нецентральности λ_s условных F-распределений тестовых статистик.

Численный расчет коэффициентов ассиметрии и эксцесса по моментам нецентрального *F*-распределения [5] показал, что даже для одного признака эти коэффициенты меньше 0,5 когда параметры нецентральности $\lambda_s > 20$. Сама дискриминантная функция имеет вид:

$$g(X) = (X - \mu^{(1)})' D^{(1)} (X - \mu^{(1)}) - (X - \mu^{(0)})' D^{(0)} (X - \mu^{(0)}) + \ln \frac{D^{(1)}}{D^{(0)}}, \qquad (11)$$

где $\mu^{(1)}$, $\mu^{(0)}$ – векторы средних по классам π_0 и π_1; $D^{(1)}$, $D^{(0)}$ – дисперсионные матрицы; $(X - \mu^{(1)})'$, $(X - \mu^{(0)})'$ – транспонированные векторы.

Дисперсионные матрицы являются диагональными, поскольку информационные признаки независимы (как функции независимых членов дисперсионного разложения (6)). Сама дискриминантная функция получена из (10) логарифмическим преобразованием:

$$g(X) = \ln \Omega$$

Практическое использование тестовых статистик. Информационная значимость статистик (7) наглядно проявляется на примере обнаружения технологических нарушений при измерительном контроле тепловых и электрических режимов маслоэкстракционных установок, используемых для получения пищевых растительных масел [6]. В таблице 1 представлены статистики $F_0^{(1)}$, $F_0^{(2)}$, $F_0^{(3)}$ для трех сигналов измерительной информации (температуры t_1 и t_2 двух камер нагрева и тока I двигателя установки), полученных для состояний π_0 (нормальная влажность сырья) и π_1 (пониженная влажность сырья).

Таблица 1 – Значения тестовых статистик для состояний π_0 и π_1

Состояние объекта контроля	Температура t_1			Температура t_2			Ток I		
	$F_0^{(1)}$	$F_0^{(2)}$	$F_0^{(3)}$	$F_0^{(1)}$	$F_0^{(2)}$	$F_0^{(3)}$	$F_0^{(1)}$	$F_0^{(2)}$	$F_0^{(3)}$
π_0	4,83	48,64	18,48	15,87	597,92	68,91	0,353	0,86	1,26
π_1	13,99	130,67	30,03	42,91	311,86	34,86	0,276	6,48	1,04

Для расчета условных средних $\mu_s^{(r)}$ и дисперсий $D_s^{(r)}$ квадратичной дискриминантной функции (11), каждая из статистик (7) должна рассматриваться как случайная величина с m_{1s} и m_{2s} степенями свободы и параметром нецентральности $\lambda_s^{(r)}$. Тогда $\mu_s^{(r)}$, $D_s^{(r)}$ и $\lambda_s^{(r)}$ будут связаны соотношениями [5]:

$$\mu_s^{(r)} = \frac{m_2}{(m_2 - 1)} \left[1 + \frac{\lambda_s^{(r)}}{m_1} \right], \qquad (12)$$

$$D_s^{(r)} = \frac{2m_2^2}{(m_2 - 2)(m_2 - 4)} \left[\frac{(1 + 2\lambda_s^{(r)})}{m_1} + \frac{(1 + \lambda_s^{(r)})^2}{(m_2 - 2)} \right]. \qquad (13)$$

Эффективность использования четного числа тестовых статистик для обнаружения технологических нарушений можно проиллюстрировать на примере двух информационных признаков:

Да

X_1 – статистика $F_0^{(3)}$ для температуры t_1;

X_2 – статистика $F_0^{(3)}$ для температуры t_2.

В таблице 2 представлены результаты расчета вероятностей ошибок α и β, а так же достоверности Р контроля технологических нарушений для одномерного (q = 1) и двумерного (q = 2) вариантов дискриминантной функции (11).

Таблица 2 – Значения вероятностей ошибок и достоверности технологического контроля

Используемые признаки	q	α	β	p
X_1	1	0,03	0,75	0,61
X_2	1	0,74	0,02	0,62
X_1 и X_2	2	0,2	0,2	0,8

Достоверность контроля (без учета априорных вероятностей состояний π_0 и π_1) ввычислялась по уравнению

$$p = 1 - (\alpha + \beta)/2 .$$

Из таблицы 2 следует, что увеличение числа тестовых статистик не только уравнивает вероятности ошибок α и β, но и повышает достоверность контроля.

Литература

1. Обнаружение изменения свойств сигналов и динамических систем: Пер. с англ./М. Басвиль, А. Вилски и др.; Под ред. М. Басвиль, А. Банвениста. – М.: Мир, 1989.–278 с.

2. Малайчук В.П., Мозговой А.В. Обработка информации в средствах и системах неразрушающего контроля.-Д.: Изд-во ДГУ, 1992.–168 с.

3. Андерсон Т. Статистический анализ временных рядов.-М.:Мир, 1975.–755 с.

4. Н. Джонсон, Ф. Лион. Статистика и планирование эксперимента в технике и науке. Методы планирования эксперимента. М.: Мир, 1981.–520 с.

5. Patnaik P.B. The noncentral χ^2 and F-distributions and their approximations.-Biometrika, 1949, vol. 36, №2, p. 202–232.

6. Щапов П.Ф., Муляров В.В. Выбор информативных показателей технологического контроля на основе двумерных моделей дисперсионного анализа// Восточно-европейский журнал передовых технологий.–3/2(15). 2005.–с. 46–49.

УДК 658.562.621(035)

Щапов П.Ф.

ВИКОРИСТАННЯ ТЕСТОВИХ СТАТИСТИК У СИСТЕМІ АВТОМАТИЧНОГО ВИМІРЮВАЛЬНОГО КОНТРОЛЮ ТЕХНОЛОГІЧНИХ ПРОЦЕСІВ

Розглянута задача коваріаційного аналізу нестаціонарних випадкових динамічних сигналів вимірювальної інформації, які використовуються для параметричного контролю технологічних процесів. Одержані аналітичні вирази для розрахунку параметрів контролю теплових та електричних режимів маслоекстракційних установок, що максимізує вірогідність контролю та поліпшує умови енергозбереження.