УДК 681.51.015

УСТОЙЧИВЫЕ ПРОЦЕДУРЫ ОБНАРУЖЕНИЯ ИЗМЕНЕНИЯ СВОЙСТВ ОБЪЕКТА В РЕАЛЬНОМ МАСШТАБЕ ВРЕМЕНИ

ВАСИЛЬЦОВА Н.В

Предлагаются процедуры оценки изменения свойств объектов. Разрабатываются робастные алгоритмы проверки нестационарности объектов. Алгоритмы и процедуры предназначены для обнаружения изменения математического ожидания и дисперсии выходной переменной объекта и могут быть использованы в реальном времени его функционирования.

Наиболее важной проблемой при решении задач идентификации объекта является обнаружение изменения его свойств в процессе функционирования. Такой подход позволяет своевременно устранить возникающие ошибки и погрешности, что влияет на качество и работоспособность создаваемых моделей.

Анализ подходов к решению подобных задач показал, что формально процедура обнаружения изменения свойств объекта может быть сведена к проверке его стационарности.

Несмотря на то, что разработано достаточно много параметрических и непараметрических критериев и алгоритмов проверки стационарности объектов [1–3], их применение в реальных условиях наталкивается на ряд сложностей, связанных с возможными случайными нарушениями хода наблюдений, ведущими к аномальным наблюдениям (выбросам). Это влечет за собой неустойчивость критерия, а следовательно, и неадекватность результатов его использования.

В работах [4,5] применяются два подхода к повышению устойчивости выводов при оценке стационарности объекта. Первый подход позволяет модифицировать критерии и сделать их "робастными" к выбросам в данных. Второй подход основан на устойчивом преобразовании данных и дальнейшем их использовании в обычных (немодифицированных) процедурах.

Еще одним препятствием эффективного использования рассматриваемых критериев является то, что большинство из них не может применяться для последовательного обнаружения изменения свойств объекта, так как осуществляет проверку гипотезы стационарности лишь по полной выборке, используя апостериорный статистический анализ.

В статье разрабатывается устойчивый к выбросам в данных алгоритм ALG, основанный на параметрических критериях проверки гипотез и используемый для оценивания стационарности объекта в реальном масштабе времени. Применение этого алгоритма для решения практических задач основано на некоторых положениях, рассмотренных ниже.

Входные и выходные переменные объекта представляют собой случайные процессы $x_i(t)$, $i=\overline{1,m}$; $y(t)$, где m – количество входных переменных. В моменты времени $t_1, t_2, t_3, \ldots, t_N, \ldots$ проводится регистрация переменных, в результате чего могут быть получены реализации случайных процессов $x_i(t)$ и $y(t)$ в виде случайных выборок $x_{i1}, x_{i2}, \ldots, x_{iN}, \ldots$ и $y_1, y_2, \ldots, y_N, \ldots$.

Реализация случайного процесса $x_i(t)$ (или $y(t)$) в общем виде является смесью следующих компонент: монотонного тренда; колебательного тренда; случайных колебаний; скачкообразных изменений переменных.

В практике построения и эксплуатации автоматизированных систем управления с идентификатором часто невозможно получить большое количество реализаций $x_i(t)$ и $y(t)$, которое необходимо для вычисления их характеристик осреднением по ансамблю. Поэтому, основываясь на понятии внутренней стационарности процессов [1], выводы об адекватности определения свойств объекта делаются по результатам анализа одной реализации. Ее проверка на стационарность осуществляется путем статистического сравнения характеристик случайного процесса, полученных на коротких интервалах времени. Если характеристики не изменяются "значимо" от интервала к интервалу, то процесс можно считать стационарным.

При отсутствии нестационарностей в случайном процессе $y(t)$ будем считать, что плотность распределения вероятностей выборки наблюдений $y_1, y_2, \ldots, y_r$ имеет вид $p(y, Q_y) = (1-\varepsilon)p_1(y, Q_{y1}) + \varepsilon\, p_2(y, Q_{y2})$, где $p_1(y, Q_{y1})$ – плотность распределения основной выборки, которая в рамках решаемой задачи считается известной (гауссовой с параметрами распределения $N(\mu, \sigma^2)$); $p_2(y, Q_{y2})$ – плотность "засоряющего" распределения, которая в общем случае неизвестна; Q_y, Q_{y1}, Q_{y2} – векторы параметров распределений; $(1-\varepsilon)$ – вероятность наблюдения выборки с плотностью распределения $p_1(y, Q_{y1})$, причем ε неизвестно, но находится в пределах $0 < \varepsilon < 1$.

Аналогично описываюся плотности распределения выборок наблюдений входных переменных x_i.

При анализе изменения свойств объекта проверяется гипотеза о слабой стационарности, т.е. о стационарности первых двух моментных характеристик распределения выборок: матемагематического ожидания и дисперсии [1].

Постановка задачи оценки стационарности объекта может быть представлена следующим образом.

Предположим, что осуществляется наблюдение входных x_i, $i = \overline{1, m}$ и выходной y переменных объекта при нормальном ходе протекания производственного процесса, причем объект может иметь несколько технологических режимов функционирования, определяемых, например, количеством перерабатываемого сырья, его составом, тепловыми и энергетическими характеристиками и т.п. Из всего множества входных переменных при решении задачи анализируются лишь те, которые определяют режим работы объекта.

При определении стационарности объекта (либо участков квазистационарности) решаются две подзадачи проверки гипотез.

Первая подзадача осуществляет последовательную проверку гипотезы отсутствия изменения технологического режима (отсутствия скачкообразного изменения математического ожидания режимной переменной) против альтернативы наличия его изменения в неизвестный момент времени t_0, т.е. в формальном виде это можно записать следующим образом. Проверяется гипотеза: $H_0 : \mu_{x_j} = \mu_0^* = \mu_a$ для $t = 1, 2, \ldots, t_0, \ldots$ против альтернативы H_1: существует такой момент времени t_0, что $\mu_{x_j} = \mu_0^*$ для $0 \le t \le t_0 - 1$ и $\mu_{x_j} = \mu_1^*$ для $t_0 \le t$, где μ_0^* и μ_1^* – математическое ожидание j–й переменной, если скачка не существует.

В реальной производственной ситуации первую подзадачу приходится решать при следующих условиях: неизвестен момент времени t_0, в который происходит скачкообразное изменение свойств; неизвестны значения параметров процесса до и после момента изменения. Поэтому результатом решения подзадачи должно явиться не только обнаружение факта изменения свойств объекта в реальном масштабе времени, но также оценка момента изменения t_0 и оценки параметров процесса до и после скачка.

Вторая подзадача заключается в том, что в рамках определенного технологического режима проверяется гипотеза существования стационарного процесса против альтернативы наличия монотонного тренда. Таким образом, в реальном масштабе времени проверяется гипотеза $H'_0 : \mu_1 = \mu_2 = \ldots = \mu_r = \ldots$ и $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_r^2 = \ldots$ против альтернативы $H'_1 : \mu_1 \neq \mu_2 \neq \ldots \neq \mu_r \neq \ldots$ и $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_r^2 = \ldots$, где $\mu_i, i = 1,2,\ldots,r,\ldots$ – математические ожидания подвыборок выходной переменной объемом $n_i \geq 2$; σ_i^2, $i = 1,2,\ldots,r,\ldots$ – дисперсии подвыборок.

Решение поставленной задачи основано на модификации двух подходов к определению изменения свойств объекта. Первый подход использует алгоритм кумулятивных сумм [6], второй – критерий отношения правдоподобия (ОП) [7].

Разработанный Е.С.Пейджем алгоритм кумулятивных сумм (КС) [6], имеющий простую вычислительную процедуру и работающий в реальном масштабе времени, основан на следующем правиле остановки:

$$t_a = \inf\left\{t : q_t \geq h_1\right\}, \qquad (1)$$

где t_a – оценка момента времени обнаружения изменения; q_t – решающая функция, причем $q_0 = 0$; $(G_t)^+ = \max(0, G_t)$; $q_t = \left\{q_{t-1} + \ln\left[p\left(x_t, \mu_1^*\right) / p\left(x_t, \mu_0^*\right)\right]\right\}^+ = \left(G_t\right)^+$; $p\left(x_t, \mu_0^*\right)$ и $p\left(x_t, \mu_1^*\right)$ – плотность распределения x до и после скачка; μ_0^*, μ_1^* – математическое ожидание x до и после скачка; h_1 – порог критерия.

Однако алгоритм КС используется в виде (1) только для известных μ_0^* и μ_1^*.

Поэтому для решения первой подзадачи используется следующая двухэтапная процедура, обнаруживающая изменение свойств объекта в реальном масштабе времени при наличии аномальных наблюдений и при минимальной информации о параметрах μ_0^* и μ_1^*. На первом этапе определяется момент обнаружения изменения свойств объекта t_a при помощи предлагаемого модифицированного алгоритма КС, который имеет следующий вид:

$$t_a = \inf\left\{t : q_t \geq h_1, \text{ если } q_{(t_a+i)} \geq h_1, \ i = \overline{1, k_a}\right\}, \qquad (2)$$

где

$$q_t = \left\{ q_{t-1} + \ln\left[p\left(x_t, \mu_{1\,t}^*\right) / p\left(x_t, \mu_{0\,t}^*\right) \right] \right\}^+ = (G)^+;$$

$$\left(G_t\right)^+ = \max\left(0, G_t\right);\ \mu_{0\,t}^* = \mu_{0\,(t\text{-}1)}^*\left(1 - \alpha^*\right) + \alpha^* x_t;$$

${\mu_{1_t}}^* = {\mu_{0_t}}^* + \Delta\mu$; α^* - постоянный коэффициент, $0 < \alpha^* < 2$; $\Delta\mu$ - минимальная величина скачка, который может быть обнаружен; задается априори, учитывая технологические условия перехода с одного режима работы на другой; k_a - количество последовательных наблюдений x_t, для которых должно выполняться условие $q_t \geq h_1$.

Робастность алгоритма (2) достигается введением в него дополнительного условия $q_{(t_a+i)} \geq h_1$, $i = \overline{1, k_a}$, позволяющего уменьшить число неверных обнаружений свойств объекта t_a.

На втором этапе определяется действительный скачок математического ожидания, который произошел в оцененный момент времени. Для этого используется критерий ОП в виде

$$q_{t_a}^* = \max \sum_{i=t_a-k_a}^{t_a} \left\{ \ln\left[p\left(x_i, \mu_{1t_a}''\right) - p\left(x_i, \mu_{0\,t_a}^*\right) \right] \right\} \geq h_2, \qquad (3)$$

где μ_{1t_a}'' – оценка максимального правдоподобия математического ожидания x_t после скачка; h_2 – порог критерия.

Для решения второй подзадачи также может быть использован критерий ОП. Он позволяет определить одновременное изменение среднего и дисперсии, что значительно уменьшает объем вычислений по сравнению с раздельным анализом этих параметров.

Для рассматриваемого в статье вида распределения выборки $y_1, y_2, \ldots, y_N$ статистика критерия ОП вычисляется как

$$\lambda = \prod_{i=1}^{r} \left(\sigma_i^2 / \sigma_0^2 \right)^{n_i/2}, \qquad (4)$$

$\sigma_i^2 = \left(1/n_i\right) \sum_{j=1}^{n_i} \left(y_{ij} - \mu_i\right)^2$; $\mu = \left(1/N\right) \sum_{i=1}^{r} n_i \mu_i$; $\sigma_0^2 = \left(1/N\right) \sum_{i=1}^{r} n_i \left[\sigma_i^2 + \left(\mu_i - \mu\right)^2 \right]$ – дисперсия выборки объема N.

Критерий (4) очень чувствителен к отклонению от нормальности, к наличию выбросов в данных и использует выборку фиксированного объема N. Существует его робастный аналог [4], однако он предполагает сложные вычислительные процедуры и требует априорного знания дисперсии выборки. В связи с этим в статье предлагается устойчивая к выбросам модификация критерия (4), позволяющая использовать его в реальном масштабе времени.

Статистику критерия ОП, используемую для решения подзадачи, представим следующим образом:

$$\lambda_i^M = \lambda_{i-1}^M \left(\sigma_{i\,n_i}^2 / \sigma_{0\,i}^2 \right)^{n_i/2}, \; i = 2,3,4,\ldots, \tag{5}$$

где $\lambda_1^M = \left(\sigma_1^2 / \sigma_0^2\right)^{n_i/2}$; σ_{ij}^2, $\sigma_{0_i}^2$ вычисляются по итерационным формулам:

$$\sigma_{ij}^2 = \mu_{ij}^{**} - \mu_{ij}^2; \quad \mu_{ij}^{**} = \mu_{j\,(j-1)}^{**}\left(1-\alpha_1\right) + \alpha_1 y_{ij}^2; \; \mu_{11}^{**} = \left(y_{ij}\right)^2;$$

$$\mu_{ij} = \mu_{i\,(j-1)}\left(1-\alpha_2\right) + y_{ij}\alpha_2; \; i = 1,2,\ldots; \quad j = \overline{1, n_i}\,; \; \widetilde{\mu}_i = \widetilde{\mu}_{i-1}\left(1-\alpha_4\right) + \alpha_4 y_{i\,n_i};$$

$$\sigma_{0_i}^2 = \sigma_{0\,(i-1)}^2\left(1-\alpha_3\right) + \alpha_3\left[\sigma_{1\,n_i}^2 + \left(\mu_{i_{n_i}} - \bar{\mu}_i\right)^2\right]; \; \sigma_{0_1}^2 = 0.$$

Коэффициенты $\alpha_i, i = \overline{1,4}$ изменяются в пределах $0 < \alpha_i < 2$.

На каждой i-й итерации статистика $\log\lambda_i^M$ сравнивается с критическим значением h_3, т.е. с величиной, распределенной по χ^2-распределению с r степенями свободы. Если $\log\lambda_i^M \chi_{r,\alpha}^2$, где α – уровень значимости критерия, то гипотеза о стационарности объекта отвергается. Устойчивость к выбросам достигается применением подхода, позволяющего преобразовывать выборку данных с последующим использованием этой выборки для проверки гипотезы стационарности по критерию (5).

В работах [8,9] рассматриваются простые и эффективные критерии отбраковки одного или нескольких аномальных наблюдений, однако они основаны на упорядочивании выборки, что значительно увеличивает время проведения расчетов.

Используемая в статье процедура не требует упорядочивания выборки, а ее минимальный и максимальный элементы определяются при последовательном сравнении текущего

$y_{(i+1),j}; i=1,2,\ldots; j=\overline{1,n_i}$ с границами $(-3\sigma_{ij}+\mu_{ij})$ и $(3\sigma_{ij}+\mu_{ij})$. Если $y_{ij} \le (-3\sigma_{ij}+\mu_{ij})$ или $y_{ij} > (-3\sigma_{ij}+\mu_{ij})$, то для наблюдения y_{ij} используется критерий обнаружения выбросов Граббса [8], статистика которого представлена как $T_y = (y_{ij} - \mu_{ij})/\sigma_{ij}$.

При $T_y < h_4$ нет оснований считать, что y_{ij} – аномальное наблюдение. Если же $T_y \ge h_4$, то y_{ij} заменяется математическим ожиданием μ_{ij} и процедура проверки продолжается. Критическое значение критерия h_4 при заданном уровне значимости α может быть найдено из таблицы в [10].

Разработанные модификации алгоритма КС (2) и критерия ОП (5) положены в основу алгоритма ALG.

Алгоритм предусматривает выполнение следующих процедур.

Шаг 1. Измерение значения режимных переменных x_t и выходной переменной y_t.

Шаг 2. Вычисление решающей функции q_t по формуле (2) и проверка условия $q_t < h_1$. При его выполнении – переход на шаг 8. При невыполнении – осуществление шагов 3 –7.

Шаг 3. Проверка наличия выбросов в x_t при помощи условия $i \ge k_a$.

Шаг 4. Определение $\mu_1^{i\prime}$ и вычисление статистики $q_{t_a}^*$ по формуле (3).

Шаг 5. Проверка условия $q_{t_a}^* \ge h_2$.

Шаг 6. Вычисление момента скачка $t_a = t - k_a$.

Шаг 7. Определение режима работы объекта.

Шаг 8. Проверка наличия выбросов в y и их исключение при помощи статистики $T_y = (y_{ij} - \mu_{ij})/\sigma_{ij}$.

Шаг 9. Формирование подвыборок данных объемом n_i.

Шаг 10. Проверка существования монотонного тренда y при помощи статистики (5). Если условие выполняется, то производится корректировка структуры и параметров модели.

Преимуществом использования алгоритма ALG является следующее. Он основан на несложных, хорошо известных с теоретической точки зрения вычислительных процедурах, легко применяемых в реальном масштабе времени. Позволяет одновременно определять монотонное изменение математического ожидания и дисперсии выходной переменной y. Алгоритм является устойчивым к выбросам в x и y.

Литература: 1. *Бендат Дж., Пирсон П.* Прикладной анализ случайных данных/ Пер. с англ. М.: Мир, 1989. 540 с. 2. *Кендел М.Г.* Временные ряды. М.: Финансы и статистика, 1981. 199 с. 3. *Рунион Г.* Справочник по непараметрической статистике. М.: Финансы и статистика, 1982. 198 с. 4. *Хьюбер П.* Робастность в статистике/ Пер. с англ. М.: Мир, 1984. 304 с. 5. Устойчивые статистические методы оценки данных/ Под ред. *Л.И. Лонера, Г.Н. Уилкинсона*/ Пер. с англ. М.: Машиностроение, 1984. 232 с. 6. *Никифоров И.В.* Последовательное обнаружение изменения свойств временных рядов. М.: Наука, 1983. 200 с. 7. *Обнаружение* изменения свойств сигналов и динамических систем/ *М. Бассвиль, А. Вилски, А. Банвенист и др.*/ Под ред. *М. Бассвиль, А. Банвениста*. М.: Мир, 1989. 278 с. 8. *Мюллер П., Нойман П., Шторл Р.* Таблицы по математической статистике. М.: Финансы и статистика, 1982. 287 с. 9. *Дэйвид Г.* Порядковые статистики. М.: Наука, 1979. 336 с. 10. *Смоляк С.А., Титаренко В.П.* Устойчивые методы оценивания: Статистическая обработка неоднородных совокупностей. М.: Статистика, 1980. 208 с.

Поступила в редколлегию 28.05.98

УДК 681.5.01

КОНЦЕПТУАЛЬНАЯ МОДЕЛЬ СИСТЕМЫ ИДЕНТИФИКАЦИИ В СОВРЕМЕННЫХ СИСТЕМАХ УПРАВЛЕНИЯ ТЕХНОЛОГИЧЕСКИМИ ОБЪЕКТАМИ

АНТОНОВ В.А., ШАМША Т.Б., РЕПКА В.Б.

Рассматривается статистический подход к оценкам эффективности алгоритмов идентификации статических характеристик в условиях мультиколлинеарности и зашумленности объектов. Приводятся основные правила автоматического выбора параметров гребневого оценивания и описывается модифицированный алгоритм онлайновой идентификации.

1. Введение

При разработке автоматизированной системы принятия решений по выбору метода идентификации динамических и статических объектов возникает проблема определения качественных показателей алгоритмов для построения базы знаний, которая позволит автоматически выбирать наиболее эффективный метод.

Известно, что практически все технологические объекты имеют зашумленность и мультиколлинеарность. В этом случае область