

**ИССЛЕДОВАНИЕ ПРОЦЕДУРЫ ОБНАРУЖЕНИЯ
РАЗЛАДКИ ВРЕМЕННОГО РЯДА НА ОСНОВЕ
ВЗВЕШЕННОГО ГОЛОСОВАНИЯ**

*Артемов Алексей Валерьевич¹
Бурнаев Евгений Владимирович²*

1: *Аспирант, физический факультет МГУ имени М. В. Ломоносова, Москва,
Россия*

2: *Зав. сектором анализа данных и моделирования, ИПФИ РАН
E-mail: artemov@physics.msu.ru, burnaev@iitp.ru*

Задача «о разладке», возникающая при динамическом анализе статистических данных в режиме реального времени, состоит в скорейшем обнаружении момента θ резкого изменения характеристик наблюдаемого случайного процесса $X = (X_t)_{t \geq 0}$, имеющего представление

$$X_t = \begin{cases} X_t^\infty, & \text{если } 0 \leq t < \theta, \\ X_t^0, & \text{если } t \geq \theta, \end{cases} \quad (1)$$

где процессы $X^\infty = (X_t^\infty)_{t \geq 0}$ и $X^0 = (X_t^0)_{t \geq 0}$ отвечают наблюдениям до и после момента θ появления разладки. Процедура обнаружения разладки процесса X в реальном времени — это момент подачи тревоги τ , минимизирующий (в заранее оговоренном усредненном смысле) время $\tau - \theta$ запаздывания в обнаружении момента θ .

Известен ряд процедур скорейшего обнаружения момента θ , оптимальных для различных моделей разладки (структуры наблюдений до и после момента θ и сценариев появления самого момента θ) [1, 2]. На практике, как правило, априорная информация о модели разладки ограничена или отсутствует, что приводит к существенному снижению эффективности процедур ее детектирования. В этом случае естественно рассматривать процедуры обнаружения (ансамбли), агрегирующие информацию о множестве «элементарных» детектирующих процедур (статистик) для принятия решения об обнаружении разладки. В настоящей работе мы исследуем подход к обнаружению разладки временного ряда, основанный на агрегации процедур скорейшего обнаружения разладки методом взвешенного голосования.

Пусть для обнаружения разладки в процессе X используются N «элементарных» процедур обнаружения Π_1, \dots, Π_N . Процедура Π_k заключается в том, что сигнал тревоги поднимается в момент τ_k

первого выхода некоторой статистики S_t^k на заданный уровень h_k : $\tau_k = \inf \{t \geq 0, S_t^k \geq h_k\}$, $k = 1, \dots, N$. Рассмотрим следующую агрегированную схему обнаружения разладки, использующую все N сигналов S_t^1, \dots, S_t^N . Пусть $a = (a_t)_{t \geq 0}$ — статистика, задаваемая соотношением

$$a_t = \sum_{k=1}^N \alpha_k \varphi(\beta_k; s_t^k), \quad (2)$$

в котором $s_t^k = S_t^k/h_k$ — масштабированное значение статистики S_t^k , α_k — вес статистики s_t^k в голосовании, $\varphi(\cdot) : [0, \infty) \rightarrow [0, 1]$ — сигмоидная функция, отображающая множество значений статистики s_t^k в отрезок $[0, 1]$ так, что значение $\varphi(\beta_k; s_t^k)$ интерпретируется как вероятность наступления разладки «по версии» статистики s_t^k , и β_k — вектор параметров функции $\varphi(\cdot)$, для статистики s_t^k , $k = 1, \dots, N$. Процедура обнаружения разладки с помощью агрегированной статистики a_t состоит в том, что сигнал тревоги подается в момент времени

$$\tau_a = \inf \{t \geq 0, a_t \geq h_a\}, \quad (3)$$

где h_a — заданное пороговое значение.

Мы предлагаем эффективный алгоритм настройки коэффициентов процедуры (2), основанный на оптимизации функции штрафа специального вида. Вычислительные эксперименты на реальных и искусственных данных показывают высокую эффективность предложенного в (2)–(3) алгоритма в случае, когда модель разладки неизвестна или задана неточно.

Литература

1. Ширяев А. Н. Вероятностно-статистические методы в теории принятия решений. М.: МЦНМО, 2011.
2. Polunchenko A. S., Sokolov G., Du W. Quickest Change-Point Detection: A Bird's Eye View. In Proceedings of 2013 Joint Statistical Meetings, August 3–5, 2013, Montreal, Quebec, Canada.