

УДК 629.7.071

ИСПОЛЬЗОВАНИЕ МЕТОДОВ КЛАСТЕРНОГО АНАЛИЗА ИЗМЕРИТЕЛЬНОЙ ИНФОРМАЦИИ ДЛЯ ОПЕРАТИВНОГО И СВОЕВРЕМЕННОГО ВЫЯВЛЕНИЯ НЕШТАТНЫХ СИТУАЦИЙ ПРИ ИСПЫТАНИЯХ ИЗДЕЛИЙ РАКЕТНО-КОСМИЧЕСКОЙ ТЕХНИКИ

Атаманчук Ю.И., начальник сектора, филиал ФГУП «ГНПРКЦ «ЦСКБ-Прогресс» - ОКБ «Спектр», 4373@spectr.ryazan.ru

Луковников К.Б., ведущий инженер-программист, филиал ФГУП «ГНПРКЦ «ЦСКБ-Прогресс» - ОКБ «Спектр», 4373@spectr.ryazan.ru

Фалеев О.В., начальник отдела, филиал ФГУП «ГНПРКЦ «ЦСКБ-Прогресс» - ОКБ «Спектр», 4373@spectr.ryazan.ru

Ключевые слова: измерительная информация, РКТ, нештатные ситуации, ТМП, кластерный анализ, контроль качества, информативность.

Предлагаемый алгоритм выявления нештатных ситуаций при испытаниях изделий ракетно-космической техники заключается в сравнении измерительной информации, характеризующей текущее состояние изделия РКТ с фрагментами измерительной информации, характерными для нештатных ситуаций и выдаче соответствующего заключения.

Модель выявления нештатных ситуаций может осуществляться различными способами и характеризоваться такими параметрами как: частота опроса датчиковой аппаратуры, тип параметра, время функционирования, наличие разрывов, монотонность и многими другими.

Алгоритм выявления нештатных ситуаций основан на априорном накоплении измерительной информации предыдущих испытаний в базе данных (БД) и последующем сопоставлении эталонных данных с анализируемыми по следующим методикам:

- 1) статистической;
- 2) корреляционной;
- 3) графическо-визуальной.

При этом вывод о наступлении нештатной ситуации носит вероятностный характер. Значение вероятности зависит от ряда факторов, таких как: номенклатура параметров признакового пространства, полнота и достаточность БД признаков используемой методики (1-3) и может колебаться в пределах 0.7-0.99. Наиболее точной, но и существенно ресурсоёмкой, является методика 3 (графическо-визуальная).

После проведения работ по анализу измерительной информации и получения истинного заключения о наступлении нештатной ситуации должна пополняться БД признаков.

Ввиду неравнозначной точности и ресурсоёмкости методик (1-3), более качественный результат даёт их последовательное использование с применением полученных результатов.

Статистическая методика предполагает для идентификации события использование статистических характеристик, оцениваемых по полученным данным. Их накопление, статистическая обработка и последующий

В настоящей статье описан алгоритм выявления нештатных ситуаций при испытаниях изделий ракетно-космической техники с использованием статистических характеристик измерительной информации для идентификации (классификации), контроля качества и определения информативности ТМП.

анализ на основе сравнения с эталонными позволяют идентифицировать изменение состояния изделия РКТ.

Пусть для i -го параметра выделенные из общего объема измерительной информации измерения образуют множество:

$$X_i = \{x_1, x_2, x_3, \dots, x_{M_i}\}, i = 1, 2, \dots, K_{II}, \quad (1)$$

где M_i - объём выборки. Совокупность всевозможных значений измерений параметра (словоформу) представим как алфавит параметра:

$$A_i = \{a_1, a_2, \dots, a_j, \dots, a_{N_i}\},$$

$$a_j > a_{j-1}, a_j \neq a_k, j, k \in \{1, 2, 3, \dots, N_i\}, i = 1, 2, \dots, K_{II}, \quad (2)$$

где a_j - словоформа алфавита $A_i \subset X_i$, в качестве которых будем использовать словарь нулей и единиц или словарь битовых словоформ.

Объединим N_i в множество N_y , упорядоченное по убыванию:

$$N_y = \{N_1, N_2, \dots, N_i, \dots, N_{K_{II}}\},$$

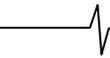
$$N_i > N_{i-1}, i \in \{1, 2, 3, \dots, K_{II}\}, \quad (3)$$

Из (3) выделим $N_{\max} = N_1$, $N_{\min} = N_{K_{II}}$ и доопределим множества разброса A_{\max} и A_{\min} мощностей алфавитов:

$$A_{\max} = \{N_{\max, l}, l = 1, 2, 3, \dots\}, A_{\min} = \{N_{\min, l}, l = 1, 2, 3, \dots\}, \quad (4)$$

где $N_{\max, l}, N_{\min, l} \in N_y$ - мощности алфавитов (3) на l -м сеансе анализа.

Если хотя бы одно их текущих значений N_{\max}, N_{\min} выходит за диапазон разброса соответствующего множества (4), то имеется изменение смыслового содержания информации. Примем решение о включении N_{\max}, N_{\min} в (4) после анализа по $v = 1, 2, \dots, V$ выборкам путём оценки диапазонов разброса мощностей их алфавитов $N_{\min, v}, N_{\max, v}$. Если для всех $v = 1, 2, \dots, V$



$$N_{\min} - \eta < N_{\min,y} < N_{\min} + \eta, N_{\max} - \mu < N_{\max,y} < N_{\max} + \mu,$$

где η и μ - допуски, то N_{\max}, N_{\min} включим в (4). Выбор значений η и μ определяется исходя из значимости изменения смыслового содержания и позволяет идентифицировать либо несущественные изменения, связанные с физикой контролируемых процессов, либо радикальные изменения, т.е. наступление факта нештатной ситуации.

Произведём валовую маркировку элементов (1), т.е. определим количество измерений i -го параметра Q_{aj} для каждого $a_j \in A_i$ в (2), и построим эмпирическую плотность распределения словоформ (монограмм):

$$P_i = \{p_j = Q_{aj}^a / M_i, j = 1, 2, \dots, N_i\}, i = 1, 2, \dots, K_{\Pi}. \quad (5)$$

Под биграммой будем понимать два рядом стоящих значения (1). Анализ биграммных зависимостей позволяет более детально и глубоко оценить информативность параметров. Для построения плотности распределения частот встречаемости биграмм произведём маркировку (1) таким образом, что каждое последующее значение биграммы берется со сдвигом на одно значение относительно начала предыдущей. В результате и получим матрицу:

$$W_i = \begin{vmatrix} w_{1,1} & w_{1,2} \dots & w_{1,N_i} \\ \dots & w_{j,k} & \dots \\ w_{N_i,1} & w_{N_i,2} \dots & w_{N_i,N_i} \end{vmatrix}, j, k \in \{1, \dots, N_i\}, i = 1, 2, \dots, K_{\Pi}, \quad (6)$$

где $w_{j,k} = Q_{aj,k} / M_i$ - частота встречаемости биграммы (a_j, a_k) .

Для проверки отличия распределения от равновероятного вычислим статистику $\chi^2_{M_i}$ для монограмм и статистику $\chi^2_{B,i}$ для биграмм /68,90,110/:

$$\chi^2_{M,i} = \sum_{j=1}^{N_i} \left(p_j - M_i / N_i \right)^2 / \left(M_i / N_i \right), i = 1, 2, \dots, K_{\Pi}, \quad (7)$$

$$\chi^2_{B,i} = \sum_{j=1}^{N_i} \sum_{k=1}^{N_i} \left(w_{jk} - (M_i - 1) / N_i^2 \right)^2 / \left((M_i - 1) / N_i^2 \right), i = 1, 2, \dots, K_{\Pi}, \quad (8)$$

где M_i / N_i и $(M_i - 1) / N_i$ - элементы равновероятного распределения.

Далее по уровню отбраковки χ^2_{sp} с доверительной вероятностью P_{Δ} проверим правильность одной из гипотез: H_0 - значения параметра распределены равновероятно; H_1 - значения параметра распределены неравновероятно. Если $\chi^2_i < \chi^2_{sp}$, то принимается H_0 , в противном случае - альтернативная гипотеза H_1 . Равновероятное распределение является важным признаком изменчивости смыслового содержания. На параметры, для которых принимается гипотеза H_0 для монограмм и биграмм одновременно, следует обращать особое внимание. Для них проведём анализ по следующим направлениям: проверка информативности; проверка на соответствие эталонным распределениям; анализ законов распределения, характеризующих непрерывность сообщений.

Проверку информативности (избыточности) будем вести на основе статистик, аналогичных энтропии. Для этого по (5) и (6) определим среднюю информативность

одного символа (монограммы) H_1 и среднюю информативность биграмм H_2 :

$$H_1 = - \sum_{j=1}^{N_i} \left(p_j / N_i \right) \log \left(r_j / N_i \right), \quad (9)$$

$$H_2 = - 1/2 \sum_{j=1}^{N_i} \sum_{k=1}^{N_i} \left(w_{jk} / N_i \right) \log \left(w_{jk} / N_i \right), \quad (10)$$

Отметим, что измерения параметров при испытаниях изделий РКТ, как правило, обладают большой избыточностью, поэтому должны выполняться условия

$$H_1 \ll H_{1,\max} = - \sum_{j=1}^{N_i} \left(1 / N_i \right) \log \left(1 / N_i \right) = \log \left(N_i \right), \quad (11)$$

$$H_2 \ll H_{2,\max} = - 1/2 \sum_{j=1}^{N_i} \sum_{k=1}^{N_i} \left(1 / N_i^2 \right) \log \left(1 / N_i^2 \right) = \log \left(N_i \right). \quad (12)$$

Из (11) и (12) вычислим соответствующие избыточности:

$$I_1 = \frac{1 - H_1}{H_{1,\max}}, I_2 = \frac{1 - H_2}{H_{2,\max}}.$$

Для малоизменчивых параметров при больших M_i должно выполняться $I_2 > I_1 \gg 0$. Изменчивые параметры, напротив, обладают нулевой избыточностью, т.е. $I \rightarrow 0$. Кроме того, для каждого разряда ИС будем также вычислять его избыточность I_l , $l = 1, \dots, P_{IC}$, с использованием более глубоких зависимостей, чем биграммные. Отметим, что, во-первых, для малоизменчивых непрерывных параметров $I_l \rightarrow 1$ для $l \rightarrow P_{IC}$, во-вторых, для изменчивых (зашумлённых) параметров $I_l \rightarrow 0$ для всех $l = 1, \dots, P_{IC}$, в-третьих, для прерывистых параметров $0 < I_l < 1$.

После принятия гипотезы H_1 по критериям (7), (8) продолжим оценку достоверности параметров путём проверки степени соответствия их эмпирических распределений эталонным. Вначале по (5) определим распределения рангов:

$$R_i = \{r_j \leftrightarrow p_j, r > r_{j+1}, j = 1, \dots, N_i\}, i = 1, 2, \dots, K_{\Pi}, \quad (13)$$

и проверим их соответствие эталонным распределениям

$$R_i^{\exists} = \{R_{i,s}^{\exists} = \{r_j^{\exists}, j = 1, 2, \dots, N_{i,s}\}, s = 1, 2, \dots, S_i\}, i = 1, 2, \dots, K_{\Pi}, \quad (14)$$

полученных при анализе предыдущих сеансов или спрогнозированных теоретически, где $N_{i,s}$ - мощность алфавита для каждого из $R_{i,s}$.

Распределения (14) содержат статистическую информацию о законе распределения вероятностей изменений параметров - средние вероятности использования словоформ (2) в последовательностях (1). При этом, можно формировать наборы вида (14) для разной информации и условий анализа. Поставим в соответствие каждому параметру наборы вида:

$$\chi_i^2, P_i \leftrightarrow \{R_{i,s}^{\exists}, N_{i,s} \in [A_{\min,s}, A_{\max,s}], s = 1, 2, \dots, S_i\}, i = 1, 2, \dots, K_{\Pi}, \quad (15)$$

где $A_{\min,s}, A_{\max,s}$ - множества мощностей алфавитов, аналогичные (4) для распределений (13). Затем вычислим:

$$\chi_{i,s}^2 = \sum_{j=1}^l \left(r_j - r_j^{\exists} \right)^2 / r_j^{\exists}, s = 1, 2, \dots, S_i, i = 1, 2, \dots, K_{\Pi}, \quad (16)$$

где $L = N_{is}$, если $N_i > N_{is}$ и $L = N_i$, если $N_{is} > N_i$.

Выберем $R_{i,s}^2$, у которого $\chi_{\min,s}^2 = \text{Min} \chi_{i,s}^2$. Если $\chi_{\min,s}^2 < \chi_{kp,s}^2$, то параметр идентифицирован. В противном случае среди (13) нет распределения, соответствующего данному параметру. Тогда либо данный параметр ранее не опрашивался, либо измерения параметра сильно отличаются от предыдущих, либо имеет место разрушение смыслового содержания помехами. Мера (16), оцениваемая по серии выборок (1), даёт оценки достоверности и качества измерительной информации.

Для построения плотностей распределения, характеризующих непрерывность изменения значений параметров (1), определим последовательность $X_{pi} = \{x_2 - x_1, x_3 - x_2, \dots, x_{M_i} - x_{M_i-1}\}$, $x_{pi} < aN_i$ и произведём маркировку положительных, отрицательных и нулевых приращений, которые формируют плотности распределений вида:

$$Q_i^+ = \{q_j^+, j = 1, 2, \dots, N_i\}, \quad Q_i^- = \{q_j^-, j = 1, 2, \dots, N_i\},$$

$$i = 1, 2, \dots, K_{II}, \quad (17)$$

где $q + j, q - j \in X_{pi}$ - количества положительных и отрицательных приращений.

Анализ (17) будем проводить по критерию, аналогичному (15), сопоставлением Q_i^+, Q_i^- с эталонными $Q_{\pm i,s}, Q_{\mp i,s}$ и с равномерно распределёнными. Кроме того, наличие в (17) большого числа максимальных приращений свидетельствует о высокой динамике поведения параметра. И напротив, большое число минимальных элементов и стремление максимальных к 0 в (16) говорит о ровной динамике параметра.

Таким образом, рассмотренный алгоритм позволяет решить задачу выявления нештатных ситуаций на семантическом уровне, т.е. идентифицировать принадлежность данных предоставленных фрагментов измерительной информации к той или иной типовой структуре, характеризующей факт наступления нештатной ситуации. Этот алгоритм включает последовательное выполнение следующих процедур. Во-первых, по выборкам (1) выполняется оценка распределения данных предоставленного фрагмента по (2), (5), (6), (13), (16), а также оценка мощностей алфавитов (3). Во-вторых, на основе сопоставления полученных оценок распределений с набором эталонных по (7), (8), (9)-(16) делается вывод о характерной структуре измерительной информации. В-третьих, для оценки степени доверия принятого решения оценивается информативность сообщения по (9)-(12).

Наполнение базы эталонов (14) может быть проведено за ограниченное время путём анализа как перекрывающихся, так и неперекрывающихся выборок анализируемых информационных процессов, а также формированием теоретических распределений. Устойчивость оценивания достигнута анализом усреднённых эмпирических распределений выборок данных и их сравнением с эталонами.

Корреляционная методика предназначена для вычисления оценок корреляционной функции по выборке данных и предусматривает расчёт автокорреляционной и взаимокорреляционной функций. Алгоритм позволяет определить оценки периодичности путем замера периода межпиковых расстояний автокорреляционной функции (АКФ), вычисляемой по выборке из анализируемой

последовательности данных.

Эта обработка используется для количественной оценки степени схожести фрагментов измерительной информации: эталонного и текущего. При этом величина статистического точечного коэффициента взаимной корреляции определяет степень соответствия фрагментов друг другу.

$$K_x(mT) = \frac{1}{(N_s - m) \sum_{i=1}^{N_s - m} [X(iT_0) - M(XiT_0)] * [X(iT_0 + mT_0) - M(X(iT_0 + mT_0))]},$$

$$m = 0, N_a - 1, \quad (18)$$

где m - номер отсчёта АКФ (mT_0 - интервал сдвига);

N_s - объём выборки (слов);

N_a - количество точек АКФ;

M^a - операция математического ожидания;

T_0 - длительность между измерениями X_i и X_{i+1} (период опроса).

При этом можно определить период или частоту собственных колебаний. Анализ характера взаимокорреляционной функции (ВКФ) используется для решения задач определения характера и свойств связи двух исследуемых процессов, эталонного и текущего. Корреляционные функции позволяют установить факты зависимости данных и оценить некоторые их характеристики.

Графическо-визуальная методика классифицирования телеметрируемых параметров (ТМП) основана на представлении измерительной информации в графическом виде и последующем экспертном оценивании степени принадлежности (схожести) графика к ранее занесенному в базу данных, характеризующему факт наступления той или иной нештатной ситуации.

Таким образом, статистические характеристики (алфавит сообщений, законы распределения словоформ и т.д.), оцениваемые для ТМП могут быть использованы для идентификации (классификации), контроля качества и определения информативности ТМП. Накопление этих характеристик, их статистическая обработка и последующий анализ на основе сравнения с эталонными значениями позволяют идентифицировать изменения смыслового содержания сообщений в результате разрушающего воздействия помех, использования кодирования или шифрования, а также дают оценки качества ТМП.

Литература

1. Эммануил С. Айфичер, Барри У. Джервис - ЦОС. Практический подход. Второе издание. Пер. с англ. Москва • Санкт-Петербург • Киев. Издательский дом «Вильямс» 2004 г.
2. Л. Рябинер, Б. Гоулд – *Теория и применение цифрой обработки сигналов*. Перевод с английского. Издательство «Мир». Москва 1978 г.
3. Фукунага К., Введение в статистическую теорию распознавания образов. Москва: Наука, 1979.

This article describes a procedure for detection of abnormal situations within rocket and space equipment tryouts using the statistical characteristics of measurement data to identify (classify), control the quality and assess the information content of telemetrical parameters.