

КЛАССИФИКАЦИЯ МНОГОПАРАМЕТРИЧЕСКИХ ОБЪЕКТОВ ПО ЭКСПЕРИМЕНТАЛЬНЫМ ДАННЫМ

Аннотация. Проведен анализ мощности классификации многомерных измерений с неизвестными статистическими закономерностям при использовании энтропийных преобразований и критериев непараметрической статистики.

Ключевые слова: выборка измерений, непараметрические критерии, энтропийные преобразования.

Постановка задачи

В настоящее время научная деятельность в технике, медицине, биологии, физики и других областях тесно связана с обработкой и анализом массивов данных, которые содержат информацию об объектах предметной области.

Рассмотрим подход к построению меры различий, которая предназначена для выявления структурных особенностей в значениях характеристик элементов исследуемых множеств, и разработке на ее основе алгоритмов классификации, например, трехпараметричных объектов по экспериментальным данным.

Цель исследования – оценка вероятностей ошибочных решений и анализ работоспособности классификации многопараметрических объектов при использовании энтропийных преобразований и комплексного критерия непараметрической статистики, в задачах неразрушающего контроля и технической диагностики.

Вычислительные эксперименты

Рассмотрим объекты классификации, находятся под влиянием трёх факторов, обуславливающих их состояние. Первая причина связана с изменением параметра $x_1(k)$ и статистически связана с измерениями $x_2(k)$ и $x_3(k)$, которые, в свою очередь, связаны законами распределения $W(x_1, x_2) = W(x_1) W\left(\frac{x_2}{x_1}\right)$.

В этом случае для выборок с нормальным распределением условный закон запишется в виде

$$W\left(\frac{x_j}{x_i}\right) = \exp - \frac{\left(\frac{x_j - a_j}{\sqrt{D_j}} - r_{ij} \frac{x_i - a_i}{\sqrt{D_i}}\right)^2}{2(1 - r_{ij}^2)} / 2\pi\sqrt{D_i D_j (1 - r_{ij}^2)}$$

Введем следующие обозначения для упрощения записи законов распределения $z_1 = \frac{x_1 - a_1}{\sqrt{D_1}}$, $z_2 = \frac{x_2 - a_2}{\sqrt{D_2}}$, $z_3 = \frac{x_3 - a_3}{\sqrt{D_3}}$, и их корреляционные зависимости $A_{11} = 1 - r_{23}^2$, $A_{22} = 1 - r_{13}^2$, $A_{33} = 1 - r_{12}^2$, $A_{12} = r_{12} - r_{13}r_{23}$, $A_{13} = r_{13} - r_{12}r_{23}$, $A_{23} = r_{23} - r_{12}r_{13}$. В этом случае случайные величины z_1 , z_2 , z_3 имеют нулевое математическое ожидание и единичную дисперсию, а их трехмерный закон распределения вероятностей запишется в виде

$$W(z_1 z_2 z_3) = \frac{\exp\left[-\frac{A_{11}z_1^2 + A_{22}z_2^2 + A_{33}z_3^2 - 2A_{12}z_1z_2 - 2A_{13}z_1z_3 - 2A_{23}z_2z_3}{2(1 - r_{12}^2 - r_{13}^2 - r_{23}^2 + 2r_{12}r_{13}r_{23})}\right]}{\sqrt{(2\pi)^3 (1 - r_{12}^2 - r_{13}^2 - r_{23}^2 + 2r_{12}r_{13}r_{23})}}$$

Условные математические ожидания и дисперсии равны $M[z_1] = 0$, $M\left[\frac{z_2}{z_1}\right] = r_{12}z_1$, $M\left[\frac{z_3}{z_1 z_2}\right] = \frac{r_{13} - r_{12}r_{23}}{1 - r_{12}} z_1 + \frac{r_{23} - r_{12}r_{13}}{1 - r_{12}} z_2$, $D[z_1] = 1$, $D\left[\frac{z_2}{z_1}\right] = 1 - r_{12}^2$, $D\left[\frac{z_3}{z_1 z_2}\right] = 1 - \frac{r_{13}^2 + r_{23}^2 - 2r_{12}r_{23}r_{13}}{1 - r_{12}^2}$.

Эти знания позволяют формировать трехмерные выборки случайных величин для проведения вычислительных экспериментов. Если $z_1 = \xi_1$, $z_2 = r_{12}\xi_1 + \sqrt{1 - r_{12}^2}\xi_2$, $z_3 = b_1 z_1 + b_2 z_2 + b_3 \xi_3$, где ξ_i – нормальные случайные величины с нулевым математическим ожиданием и единичной дисперсией ($i = 1, 2, 3$). Коэффициенты b_1 , b_2 , b_3 определим из системы уравнений

$$\begin{cases} r_{13} = b_1 + b_2 r_{12}, \\ r_{23} = b_1 r_{12} + b_2, \\ 1 = b_1^2 + b_2^2 + b_3^2 + 2b_1 b_2 r_{12}. \end{cases} \quad (1)$$

Решив систему уравнений (1) получим коэффициенты

$$b_1 = \frac{r_{13} - r_{12}r_{23}}{1 - r_{12}^2}, \quad b_2 = \frac{r_{23} - r_{12}r_{13}}{1 - r_{12}^2},$$

$$b_3^2 = 1 - \left[\left[\frac{r_{12}(r_{13} - r_{23}r_{12})}{1 - r_{12}^2} + \frac{r_{23} - r_{13}r_{12}}{1 - r_{12}^2} \right]^2 + \left[\frac{r_{13} - r_{12}r_{23}}{\sqrt{1 - r_{12}^2}} \right]^2 \right].$$

Имея в распоряжении некоторое число трехпараметрической объектов со случайными параметрами и выборками их измерений, необходимо выделить классы объектов с одинаковыми параметрами. Эту задачу можно решить путем энтропийных преобразований выборок измерений и проверкой гипотез о равенстве их оползней, масштабов и корреляционных связей на основе критериев Ван-дер-Вардена, Клотца и Буша-Винда.

Рассмотрим эту задачу на примере следующих двух видов разницы объектов. Первый вид различий: 1) объекты отличаются лишь параметром сдвига; 2) объекты отличаются лишь параметром масштаба; 3) объекты отличаются только статистическими связями (коэффициентом корреляции).

Второй вид различий (сдвиг, масштабов и корреляции): 1) объекты отличаются одним (неизвестным) параметром; 2) объекты отличаются двумя (неизвестным) параметрами; 3) объекты отличаются всеми (тремя) параметрами.

Проведем учебный вычислительный эксперимент, чтобы получить информацию о различиях соответствующих энтропийных преобразований, их гистограммы, минимальные, максимальные и средние значения, выборочные дисперсии и коэффициенты вариации и размахов. Если коэффициент вариации - это отношение $k_v = \sqrt{D^*} / M^*$, то коэффициент размаха - $k_p = (L_{\max} - L_{\min}) / \sqrt{D^*}$. Считая параметры трех эталонных гауссовых выборок измерений известными (см. таблицу 1) определим их идеальный энтропийный преобразователь $L_s(x_1(k), x_2(k), x_3(k))$.

Параметры эталонного энтропийного преобразователя

Параметры	a_1	a_2	a_3	D_1	D_2	D_3	r_{12}	r_{13}	r_{23}
Эталон	10	20	30	1	2	3	0,8	0,8	0,8

$$L_1(x_1(k), x_2(k), x_3(k)) = \frac{1}{2} \ln \left[(2\pi)^3 (1 - r_{12}^2 - r_{13}^2 - r_{23}^2 + 2r_{12}r_{13}r_{23}) D_{11}^* D_{12}^* D_{13}^* \right] + \left(\frac{A_{11}x_1^2 + A_{22}x_2^2 + A_{33}x_3^2 - 2A_{12}x_1x_2 - 2A_{13}x_1x_3 - 2A_{23}x_2x_3}{2(1 - r_{12}^2 - r_{13}^2 - r_{23}^2 + 2r_{12}r_{13}r_{23})} \right) \quad (2)$$

где $x_1(k) = a_{11} + \sqrt{D_{11}}z_1$, $x_2(k) = a_{12} + \sqrt{D_{12}}z_2$, $x_3(k) = a_{13} + \sqrt{D_{13}}z_3$.

Имея в распоряжении три выборки измерений независимых нормальных случайных величин с нулевым математическим ожиданием и единичной дисперсией $\xi_1(k)$, $\xi_2(k)$, $\xi_3(k)$, сформируем три выборки взаимокоррелированных случайных величин $z_1(k) = \xi_1(k)$, $z_2(k) = r_{12}\xi_1(k) + \sqrt{1 - r_{12}^2}\xi_2(k)$, $z_3(k) = b_1z_1(k) + b_2z_2(k) + b_3\xi_3(k)$. Изменяя их параметры исследуем статистические закономерности энтропийных преобразователей в соответствии с указанными выше типов различий смоделированных выборок от идеала. По первым видом имеем три варианта различий: 1) $a_1 = 11$, $a_2 = 21$, $a_3 = 31$; 2) $D_1 = 1,5$, $D_2 = 2,5$, $D_3 = 3,5$; 3) $r_{12} = 0$, $r_{23} = 0$, $r_{13} = 0$. Объем измерений составляет $n = 1000$.

Статистические показатели энтропийного преобразователя представим в виде таблицы 2.

Таблица 2

Статистические показатели энтропийного преобразователя

	\bar{L}^*	$\sqrt{D^*[L]}$	L_{\max}^*	L_{\min}^*	k_e	k_p
Эталон	3330	5,834	3330	3289	0,018	7,773
1	3337	5,889	3342	3294	0,017	7,7
2	3422	6,112	3446	3399	0,017	7,7
3	5,141	1,487	3,653	11,281	0,237	6,252

По данным таблицы 2 можно утверждать, что, несмотря на изменения математического ожидания и дисперсии входных данных, смещение и масштаб энтропийного преобразователя остаются практически неизменными, а при изменении корреляции происходит существенное изменение математического ожидания, и уменьшается дисперсия преобразователя. На рисунке 1 изображены гистограммы трехмерных энтропийных преобразований эталона (рисунок 1а и случая, когда входные данные отличаются смещением при неизменных масштабах и коэффициентах корреляции эталона (рисунок 1б)

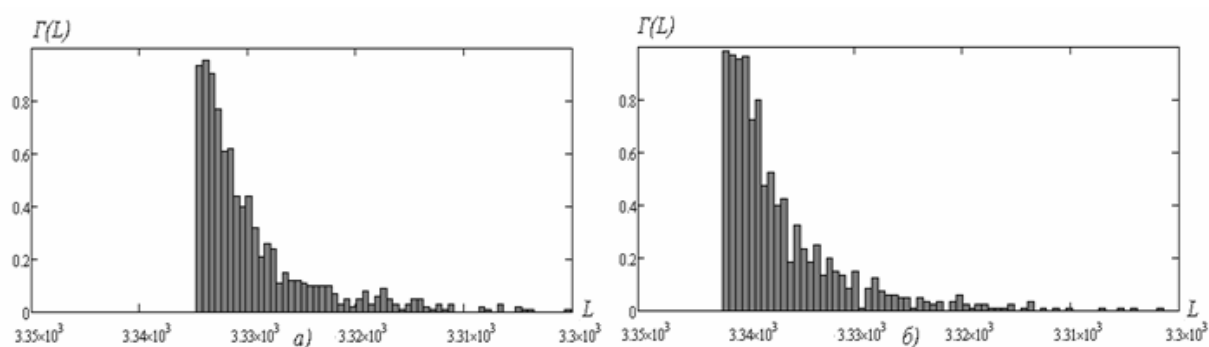


Рисунок 1 - Гистограммы трехмерных энтропийных преобразований а) эталонный энтропийный преобразователь; б) энтропийный преобразователь при уменьшении математического ожидания входных данных

Из анализа рисунка 1 следует, что вид закона распределения вероятностей энтропийного преобразователя не зависит от изменения статистических показателей входных данных.

Исследуем влияние на преобразователь второго вида различий, а именно: изменим параметры только первой из трех эталонных выборок (см. таблицу 1): 1) $a_1 = 11$, $D_1 = 1,5$, $r_{12} = 0,5$; 2) первой и второй эталонных выборок $a_1 = 11$, $D_1 = 1,5$, $r_{12} = 0,5$; $a_2 = 21$, $D_2 = 2,5$, $r_{23} = 0,7$; 3) всех трех выборок.

Статистические показатели энтропийного преобразователя представим в виде таблицы 3.

Статистические показатели энтропийного преобразователя

	\bar{L}^*	$\sqrt{D^* [L]}$	L_{\max}^*	L_{\min}^*	k_e	k_p
Эталон	3330	5,834	3330	3289	0,018	7,773
1	4769	7,323	4776	4707	0,015	9,312
2	2805	5,05	2809	2771	0,018	7,58
3	2638	4,413	2642	2609	0,016	7,363

Анализируя данные таблицы 3 приходим к выводу, что разница одного параметра влияет на статистические показатели энтропийного преобразователя намного сильнее, чем общее увеличение или уменьшение математического ожидания или дисперсии.

Для проверки возможности использования критерия Буша-Винда для выявления различий статистических закономерностей энтропийных преобразователей исследуем вероятности принятия решений по одинаковости объектов контроля по сравнению с эталоном, при условии, что второй объект имеет отличие от эталона первого и второго класса, при параметрах указанных выше. Объем измерений составляет $n = 1000$ количество реализаций 1000 Результаты опыта представлены в таблице 4.

Таблица 4

Вероятности принятия решений об однородности энтропийных преобразователей по критерию Буша-Винда

Эталон	1 класс отличий (изменяемые параметры)			2 класс отличий (изменяемые параметры)		
	МО	Дис.	Кор.	x_1	x_1 и x_2	x_1, x_2, x_3
0,994	0,003	0,007	0,894	0,793	0,433	0,004

По данным таблицы 4 можно сделать вывод о высокой эффективности совместного использования комплексного непараметрического критерия Буша-Винда и энтропийных преобразований в задачах дефектоскопии. Поскольку критерий в совокупности с энтропийными преобразованиями позволяют определить изменения как только один из параметров измерений, так и изменение статистических показателей одной выборки многомерного объекта неразрушающего контроля.

Выводы

Использование комплексного критерия непараметрической статистики Буша-Винда совокупно с энтропийными преобразованиями позволяет с высокой степенью достоверности (0,95) классифицировать объекты неразрушающего контроля. Эти методы позволяют делить изделия не только на нормальные и дефектные, но и устанавливать чем именно каждое изделие отличается от «нормы». Установленные таким образом причинно-следственные связи между измерительными параметрами объектов контроля и их отношением к определенному классу, позволяют корректировать технологию производства.

ЛИТЕРАТУРА

1. Куренков Н. И. Энтропийный подход к решению задач классификации многомерных данных / Н. И. Куренков, С. Н. Ананьев // Ежемесячный теоретический и прикладной научно-технический журнал «Информационные технологии». – М. : «Новые технологии», 2006. – № 8. – С. 50-55.
2. Jenssen R. An Information Theoretic Approach to Machine Learning : Diss. for the Deg. of Dr. Scientiarum / R. Jenssen ; Department of Physics University of Tromso. – Tromso, 2005. – 179 p.