УДК 620.179

А.И. Федорович

ИССЛЕДОВАНИЕ ПОТЕНЦИАЛЬНЫХ ВОЗМОЖНОСТЕЙ КЛАССИФИКАЦИИ ЭНТРОПИЙНЫХ ВЫБОРОК СЛУЧАЙНЫХ ВЕЛИЧИН

Аннотация. Проведен анализ возможности классификации многопараметрических объектов с использованием энтропийных преобразований. Исследовано влияние объема исходных данных на достоверность принятия решений о классе объекта контроля, а также влияние на ошибки априорных знаний об анализируемом объекте.

Ключевые слова: выборка измерений, энтропийные преобразования, классификация объектов, эффективность распознавания.

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 возможности классификации множества многопараметрических объектов, который предназначена для выявления структурных особенностей в значениях характеристик элементов исследуемых множеств.

Цель исследования – анализ эталонных выборок энтропийных преобразований и оценка работоспособности метода классификации многопараметрических объектов по экспериментальным измерениям.

Вычислительные эксперименты

Задачу оценки потенциальных возможностей классификации можно решить путем проведения вычислительных экспериментов.

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 случай-

© Федорович А.И., 2018

ные величины z_1, z_2, z_3 имеют нулевое математическое ожидание и единичную дисперсию, а их трехмерный закон распределения вероятностей запишется в виде

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

Условные математические ожидания и дисперсии равны

$$M[z_1] = 0, \qquad M\left[z_2 / z_1\right] = r_{12}z_1, \qquad M\left[z_3 / z_1z_2\right] = \frac{r_{13} - r_{12}r_{23}}{1 - r_{12}}z_1 + \frac{r_{23} - r_{12}r_{13}}{1 - r_{12}}z_2,$$

$$D[z_1] = 1, \; D\left[z_2 / z_1\right] = 1 - r_{12}^2, \; D\left[z_3 / z_1z_2\right] = 1 - \frac{r_{13}^2 + r_{23}^2 - 2r_{12}r_{23}r_{13}}{1 - r_{12}^2}.$$

Эти знания позволяют формировать трехмерные выборки случайных величин для проведения вычислительных экспериментов. Если $z_1 = \xi_1$, $z_2 = r_{12}\xi_1 + \sqrt{1 - r_{12}^2}\xi_2$, $z_3 = b_1z_1 + b_2z_2 + b_3\xi_3$, 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

$$\begin{cases} r_{13} = b_1 + b_2r_{12}, \\ r_{23} = b_1r_{12} + b_2, \\ 1 = b_1^2 + b_2^2 + b_3^2 + 2b_1b_2r_{12}. \end{cases} \qquad (1)$$

Решив систему уравнений (1) получим коэффициенты

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

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

Проведем учебный вычислительный эксперимент, чтобы получить информацию о различиях соответствующих энтропийных преобразований, их гистограммы, минимальные, максимальные и средние значения (M^*), выборочные дисперсии (D^*) и коэффициенты вариации и размахов. Если коэффициент вариации - это отношение $k_в = \sqrt{D^*} / M^*$, то коэффициент размаха – $k_р = \left(L_{\max} - L_{\min}\right) / \sqrt{D^*}$. Счи-

тая параметры трех эталонных гауссовых выборок измерений известными (см. таблицу 1) определим их идеальный энтропийный преобразователь $L_э\left(x_1(k), x_2(k), x_3(k)\right)$.

Таблица 1

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

$$L_1\left(x_1(k), x_2(k), x_3(k)\right) = \frac{1}{2}\ln\left[(2\pi)^3\left(1 - r_{12}^2 - r_{13}^2 - r_{23}^2 + 2r_{12}r_{13}r_{23}\right)D_{11}^* D_{12}^* D_{13}^*\right] + \\ + \left(\frac{A_{11}x_1^2 + A_{22}x_2^2 + A_{33}x_3^2 - 2A_{12}x_1x_2 - 2A_{13}x_1x_3 - 2A_{23}x_3x_2}{2\left(1 - r_{12}^2 - r_{13}^2 - r_{23}^2 + 2r_{12}r_{13}r_{23}\right)}\right) \quad (2)$$

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

Используя критерий сравнения $W(m)$ выделим из нее те, которые относятся к первому классу

$$R^*\left(m/1\right) = \operatorname{sgn}\left(W_0 - W(m)\right), \quad (3)$$

где $\operatorname{sgn}(x)$ – функция единичного скачка; W_0 – пороговое значение критерия сравнения. В качестве критерия сравнения, можно использовать комплексный критерий непараметрической статистики – Буша-Винда.

Очевидно, что их относительное число может служить оценкой технологии производства этих объектов

$$P_{11}^* = \frac{1}{M}\sum_{m=1}^{M}\operatorname{sgn}\left(W_0 - W(m)\right). \quad (4)$$

Сформируем по двадцать трехпараметрических выборок измерений с объемом n, имея в распоряжении параметры $(a_{11}, a_{12}, a_{13}, D_{11}, D_{12}, D_{13}, r_{12}, r_{13}, r_{23}$ і $a_{21}\ a_{22}, a_{23}, D_{21}, D_{22}, D_{23}, r_{21}, r_{31}, r_{32})$. Выберем три первых из них и оценим их параметры, например a_{11}^*, a_{12}^*, a_{13}^*, D_{11}^*, D_{12}^*, D_{13}^*, r_{12}^*, r_{13}^*, r_{23}^* и сформируем энтропийный преобразователь объектов первой класса по формуле (2).

Эксперимент 1. Цель эксперимента – исследование эталонных выборок энтропийных преобразований. В этом случае предполагается, что эталонный преобразователь сформирован с параметрами a_{11} a_{12},

a_{13}, D_{11}, D_{12}, D_{13}, r_{12}, r_{13}, r_{23}, создаются достаточно длинные эталонные выборки энтропийного преобразования, строится гистограмма и оценки статистических показателей для случаев энтропийных преобразователей по известным параметрам измерительных выборок и оценками этих параметров.

В результате проведения этого эксперимента получено гистограмму энтропийного преобразователя (2) ($n = 2000$),которая изображена на рисунке 1, при значениях параметров $a_{11} = 0$, $a_{12} = 0$, $a_{13} = 0$, $D_{11} = 1$, $D_{12} = 1$, $D_{13} = 1$, $r_{12} = 0$, $r_{13} = 0$, $r_{23} = 0$.

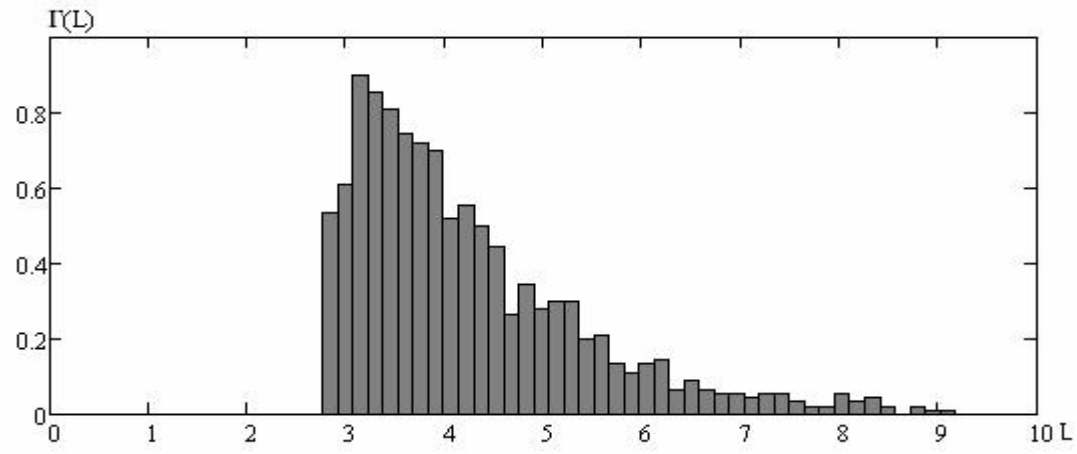

Рисунок 1 - Гистограмма энтропийного преобразователя ($\overline{L}^* = 4,245$, $\sqrt{D^*[L]} = 1,22$, $L_{\min} = 2,763$, $L_{\max} = 10,385$, $k_v = 0,287$, $k_r = 6,252$)

Проверим выбранное значение подав на энтропийный преобразователь (2) двадцать трехмерных нормальных выборок с параметрами $a_{11} = 0$, $a_{12} = 0$, $a_{13} = 0$, $D_{11} = 1$, $D_{12} = 1$, $D_{13} = 1$, $r_{12} = 0$, $r_{13} = 0$, $r_{23} = 0$, длиной $n = 50$, и оценим по формуле (3) эффективность распознавания классов объекта контроля по сложившемуся правилу. В результате получим вероятность $P \geq 0,949$, не противоречащей выбранной вероятности принятия правильных решений.

Для проверки экспериментального преобразователя воспользуемся теми же условиями, но вероятность принятия правильного решения относительно класса объекта контроля оценивать по формуле (4). В этом случае получим $P^* \geq 0,926$. То есть отсутствие информации о параметрах входных данных увеличивает вероятность принятия ошибочного решения относительно класса объекта контроля на 2,3%.

Эксперимент 2. Цель – оценка работоспособности метода классификации многопараметрических объектов по экспериментальным измерениям.

Сформируем аналогично формуле (2) преобразователи для второго и третьего классов случайных величин и с их помощью классифицируем двадцать объектов контроля каждого из указанных классов. Объекты контроля описываются трехмерными выборками нормальных случайных величин. Параметры первого класса: $a_{11} = 0$, $a_{12} = 0$, $a_{13} = 0$, $D_{11} = 1$, $D_{12} = 1$, $r_{12} = 0,7$, $r_{13} = 0,8$, $r_{23} = 0,9$. Параметры второго класса: $a_{21} = 0$, $a_{22} = 0$, $a_{23} = 0$, $D_{21} = 1$, $D_{22} = 1$, $D_{23} = 1$, $r_{12} = 0$, $r_{13} = 0$, $r_{23} = 0$, есть разница между первым и вторым классами заключается в наличии или отсутствии корреляции между измерительными параметрами. Параметры третьего класса: $a_{31} = 1$, $a_{32} = 1$, $a_{33} = 1$, $D_{31} = 2$, $D_{32} = 2$, $D_{33} = 2$, $r_{12} = 0,9$, $r_{13} = 0,7$, $r_{23} = 0,8$. Изделия, попавшие в третий класс существенно отличаются от изделий первого и второго классов по всем показателям.

Результат классификации представим в виде таблицы 2. Поскольку в этом вычислительном эксперименте речь идет о потенциальных возможностях классификации, то объем выборок измерений исследуемых выберем $n = 2000$, а параметры, формирующие энтропийные преобразователи, считать известными. Итак, согласно классификации с помощью энтропийных преобразований имеем:

Таблица 2

Количество объектов каждого класса

	Класс 1	Класс 2	Класс 3
Энтропийный преобразователь 1 класса	19	1	0
Энтропийный преобразователь 2 класса	2	17	1
Энтропийный преобразователь 3 класса	1	2	17

По данным таблицы 2 можно утверждать, что предложенный метод классификации работоспособен. При его использовании ошибки второго рода не превышают 7%. Это дает право рекомендовать его для использования при решении задач дефектоскопии.

Оценим влияние ограничения объема измерений и неизвестность точных значений параметров измерений на эффективность принятия решений контроля. Для этого по указанным в вычислительном эксперименте 1 параметрами объекта контроля сформируем выборки измерений объемом $n = 25$, $n = 50$, $n = 100$. Данные эксперимента сведены в таблице 3.

Таблица 3

Вероятности принятия правильных решений
о классе «нормы» объекта контроля

Вероятность вычисленная по точным параметрам объекта контроля			Вероятность вычисленная по оценкам параметров объекта контроля		
n=25	n=50	n=100	n=25	n=50	n=100
0,902	0,978	0,989	0,873	0,972	0,981

По данным таблицы 3 можно сделать вывод, что при объеме выборок измерений объектов контроля $n \le 25$, вероятность принятия правильного решения уменьшается, а при объемах измерений $n \ge 50$ отсутствие точной информации относительно значений параметров входных данных практически не влияет на качество классификации по энтропийным преобразователем трехмерных случайных величин.

Выводы

В задачах классификации множества объектов контроля, при отсутствии выборок измерений эталонных образцов, использование энтропийных преобразований позволит не только повысить достоверность принимаемых решений, но и проводить сравнительный анализ измерений контролируемых параметров, и оценить их причинно-следственные связи. Кроме этого, проведенный анализ объема измерений, дает возможность оценить достоверность принимаемых решений. При этом построенные на основе рассмотренных методов решающие правила могут адаптироваться и уточняться в процессе проведения неразрушающего контроля и накопления данных об объектах, контролируемых.

ЛИТЕРАТУРА

1. Кобзарь А.И. Прикладная математическая статистика/ А.И. Кобзарь. – М.: ФИЗМАТ ЛИТ, 2006. – 816 с.

2. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. – М.: Наука, 1983. – 416 с.

3. 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.

4. 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.

5. Xu Rui. Survey of clustering algorithms / Rui Xu, D. Wunsch II // IEEE Transactions on Neural Networks. – 2005. – V. 16, № 3. – P. 645.

6. Бабак В. П. Теоретические основы информационно-измерительных систем: Учебник / В. П. Бабак, С. В. Бабак, В. С. Ерёменко. – К. : ТОВ «Софія-А», 2014. – 832 с.

7. Fedorovich A. Classification of facilities multi parameters experimental measurements of their parameters / A. Fedorovich // European science review. – 2015. – № 7-8 July-August. – P. 140-142.