

ФАКТОРНЫЙ АНАЛИЗ КЛАССИФИКАЦИИ ЭНТРОПИЙНЫХ ВЫБОРОК СЛУЧАЙНЫХ ВЕЛИЧИН

Аннотация. Проведен факторный анализ влияния параметров выборок измерений при классификации многомерных объектов. Исследование проведено на основании вычислительных экспериментов с использованием энтропийных преобразований. Определено влияние сдвига, масштаба и корреляции исходных данных на достоверность принятия решений о классе объекта контроля.

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.

Для использования многопараметрических энтропийных преобразователей в задачах дефектоскопии необходимо исследовать влияние различных факторов на качество классификации. По этому поводу целесообразно провести факторный анализ и установить наиболее значимые параметры.

Цель исследования – факторный анализ влияния различных параметров исходных данных на работоспособности метода классификации многопараметрических объектов по экспериментальным измерениям. При использовании энтропийных преобразований последних.

Вычислительные эксперименты

Задачу факторного анализа информативности параметров многомерных измерений при классификации объектов, можно решить путем

проведения вычислительных экспериментов. Для этого необходимо сформировать многомерные выборки измерений.

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

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

Его энтропийное преобразование будет иметь вид:

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

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

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

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

где $\text{sgn}(x)$ – функция единичного скачка; W_0 – пороговое значение критерия сравнения

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

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

Для использования многопараметрических энтропийных преобразователей в задачах дефектоскопии необходимо исследовать влияние различных факторов на качество классификации. По этому поводу целесообразно провести факторный анализ и установить наиболее значимые параметры.

Данные для проведения вычислительных экспериментов приведены в таблице 1. эталоны приняты объект O_1 (1 класс), объект O_2 (2 класс), объект O_9 (3 класс).

Таблица 1

Параметры объектов, поступающих на контроль

	a_{11}	a_{12}	a_{13}	D_{11}	D_{12}	D_{13}	r_{12}	r_{13}	r_{23}
O_1	0	0	0	1	1	1	0,7	0,8	0,9
O_2	0	0	0	1	1	1	0	0	0
O_3	1	0	0	2	1	1	0	0	0,9
O_4	0	1	0	1	2	1	0	0,8	0
O_5	0	0	1	1	1	2	0,7	0	0
O_6	1	1	0	2	2	1	0,5	0,5	0,9
O_7	1	0	1	2	1	2	0,5	0,8	0,5
O_8	0	1	1	1	2	2	0,7	0,5	0,5
O_9	1	1	1	2	2	2	0,9	0,7	0,8

Как было отмечено выше, измеряемый параметр является носителем информации о состоянии объекта контроля и при этом воплощает в своих статистических показателях последствия той или иной причины в различии исследуемых объектов. Исходя из этого данные для проведения вычислительных экспериментов факторного анализа построены таким образом, чтобы каждый объект отличался от принятого за «норму» статистическими показателями параметра $x_1(k)$, $x_2(k)$, $x_3(k)$ или их комбинациями. Отдельно исследован вопрос о разрушении корреляции. Поскольку наличие статистической связи между параметрами объекта контроля существенно влияет на вид гистограммы энтропийного преобразователя и его среднее значение.

Во время эксперимента оценивалась вероятность идентификации исследуемого объекта контроля как принадлежащего классу объектов O_1 , O_2 или O_9 . Вероятность принятия решений исчислялась по формулам (3) и (4) в соответствии с случае когда энтропийный преобразователь был сформирован по точным значениям параметров объекта контроля и для случая, когда в их качестве выступали оценки этих значений по экспе-

риментальным данным. Результаты эксперимента сведены в таблицу 2. Количество реализаций составляет 1000.

Таблица 2

Вероятность принятия гипотезы о классе объекта контроля (расчеты по точным значениям параметров) ($n = 50$)

Объект контроля	O_1	O_2	O_3	O_5	O_6	O_7	O_8	O_9
$P\left(\frac{O_j}{O_1}\right)$	0,962	1	0,43	0,468	0,612	0,673	0,654	0,916
$P\left(\frac{O_j}{O_2}\right)$	0,991	0,997	0,57	0,412	0,589	0,573	0,645	0,897
$P\left(\frac{O_j}{O_9}\right)$	0,901	0,857	0,489	0,456	0,634	0,645	0,873	0,989

Таблица 3

Вероятность принятия гипотезы о классе объекта контроля (расчеты по оценкам значений параметров) ($n = 50$)

Объект контроля	O_1	O_2	O_3	O_4	O_5	O_6	O_7	O_8	O_9
$P^*\left(\frac{O_j}{O_1}\right)$	0,911	1	0,257	0,302	0,315	0,498	0,511	0,493	0,887
$P^*\left(\frac{O_j}{O_2}\right)$	0,903	0,945	0,263	0,314	0,298	0,487	0,521	0,479	0,891
$P^*\left(\frac{O_j}{O_9}\right)$	0,893	0,713	0,433	0,493	0,302	0,501	0,574	0,799	0,901

Из анализа данных таблицы 3 следует, что наиболее влиятельным параметром на вывод о состоянии объекта контроля является корреляционная связь. Отличие математического ожидания и дисперсии влияет на результаты незначительно. Поэтому при использовании этого метода на практике возможно исследовать малейшие отклонения корреляции измеряемых параметров. Это дает возможность устанавливать причинно-следственные связи, но вместе с тем, этот метод требует проведения детальных исследований по различия между математическими ожиданиями и дисперсиями.

Сравнение таблицы 2 и 3 позволяет оценить влияние недостатка априорной информации при формировании энтропийных преобразова-

телей. Согласно указанным вероятностям можно утверждать, что отсутствие информации о статистических показателях измерительных параметров уменьшает вероятность принятия правильных решений в среднем на 5,1%. Но вместе с тем, предложенный метод классификации позволяет уточнять параметры преобразователей при накоплении информации о статистических показателях объектов неразрушающего контроля.

Если данные таблиц 2 и 3 интерпретировать в количестве изделий, отнесенных к каждому из трех классов, при условии, что объектов с параметрами указанными в таблице 1 было по двадцать каждого вида, то получим следующие результаты сведенные в таблице 4.

Таблица 4

Вероятность принятия гипотезы о классе объекта контроля (расчеты по точным значениям параметров) ($n = 50$)

Объект контроля	O_1	O_2	O_3	O_4	O_5	O_6	O_7	O_8	O_9
Класс 1	29	28	13	15	13	17	19	18	28
Класс 2	28	29	16	19	12	16	17	18	25
Класс 3	26	24	15	14	13	18	18	24	28

По данным таблицы 4 можно сделать вывод о работоспособности предложенного метода классификации объектов неразрушающего контроля. Но наличие ошибок и, как было отмечено выше, не большая чувствительность к изменениям только оползней или масштабов в статистических показателях измерительных объектов, требует совершенствования этого метода путем объединения классификации по энтропийными преобразованиями с комплексным критерием непараметрической статистики Буша-Винда.

Выводы

Установлено, что причинно-следственные связи между контролируемым объектом и его параметрами, влияют на корреляционные зависимости измерений и находят отражение в их энтропийных преобразованиях. А также оказывают влияние на решения, принимаемые по энтропийным преобразованиям исходных данных. Этот факт позволяет

уменьшать количество ошибочных решений, принимаемых при классификации объектов контроля.

Путём энтропийного преобразования многомерные выборки трансформируются в одномерные, что позволяет при их оценки использовать стандартные критерии непараметрической статистики.

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.

REFERENCES

1. Kobzar A.Y. Prykladnaia matematycheskaia statystyka/ A.Y. Kobzar. – M.: FYZMAT LYT, 2006. – 816 s.

2. Bolshev L.N., Smyrnov N.V. Tablytsy matematycheskoï statystyky. – M.: Nauka, 1983. – 416 s.

3. Kurenkov N. Y. Entropyinyi podkhod k reshenyiu zadach klassyfykatsyy mnohomernykh dannykh / N. Y. Kurenkov, S. N. Ananev // Ezhemesiachnyi teoretycheskyi y prykladnoi nauchno-tekhnycheskyi zhurnal

«Ynformatsy-onnye tekhnolohyy». – M. : «Новые tekhnolohyy», 2006. – № 8. – S. 50-55.

4. Jenssen R. An Information Theoretic Approach to Machine Learning : Diss. for the Deg. of Dh. 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. Babak V. P. Teoretycheskye osnovy ynformatsyonno-ymerytelnykh system: Uchebnyk / V. P. Babak, S. V. Babak, V. S. Erëmenko. – K. : TOV «Sofiia-A», 2014. – 832 s.

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.