УДК 681. 518. 54

© Н. А. Лысенко

Дніпропетровський національний університет

ФОРМИРОВАНИЕ РЕШАЮЩИХ ПРАВИЛ НЕРАЗРУШАЮЩЕГО КОНТРОЛЯ НА ОСНОВЕ МЕТОДА ГРУППОВОГО УЧЕТА АРГУМЕНТОВ

Досліджується ефективність емпіричних вирішувальних правил, побудованих на основі методу групового врахування аргументів. Шляхом проведення обчислювальних експериментів проведено порівняння емпіричних вирішувальних правил з оптимальними. Отримано оцінки ймовірностей розпізнавання і досліджено їхню залежність від довжини висхідних вибірок. Розглядається питання поділу вибірок на навчальні та перевіркові.

ВВЕДЕНИЕ

Ответственным этапом любой технологии неразрушающего контроля является построение решающих правил контроля [4]. Рассмотрим задачу распознавания контролируемых объектов, если в результате предварительных исследований получены измерения информативных параметров, характеризующих нормальное и дефектное состояние объектов контроля. Такая задача может возникнуть при контроле различных материалов, конструкций и изделий ракетно-космической техники.

Существуют различные пути решения поставленной задачи. В работе [5] предлагается формирование решающих правил контроля двумя способами:

1) на основе законов распределения, восстановленных методом сглаженных дельта-функций;

2) путем оценки логарифма отношения функций правдоподобия с использованием метода группового учета аргументов.

Насколько эффективными являются эмпирические решающие правила контроля, построенные на основе метода группового учета аргументов, в сравнении с оптимальными решающими правилами? Как влияет на эффективность эмпирических решающих правил контроля объем исходных данных, т. е. длина выборок измерений? Как оценивать эффективность решающих правил, делить или не делить измерения на обучающие и проверяющие? Ответов и практических рекомендаций на все эти вопросы пока нет.

ФОРМУЛИРОВКА ЦЕЛЕЙ

Пусть объекты в норме и при наличии брака характеризуются несколькими параметрами y_1, y_2, ..., y_m, которые могут иметь различную физическую природу и размерность. Их условные многомерные законы распределения $W(|y|/H)$ и $W(|y|/Б)$ неизвестны, но заданы выборки измерений параметров объектов в норме ($y_1^{(1)}$, $y_2^{(1)}$, ..., $y_m^{(1)}$) и при наличии дефектов ($y_1^{(2)}$, $y_2^{(2)}$, ..., $y_m^{(2)}$). Исходные данные сведены в табл. 1.

По этим экспериментальным данным необходимо построить эмпирические решающие правила распознавания на основе метода группового учета аргументов, оценить эффективность полученных решающих правил и сравнить их с оптимальными, исследовать зависимость эффективности эмпирических решающих правил от длины обучающих выборок и изучить вопрос о целесообразности деления исходных выборок измерений на обучающие и проверяющие.

ОСНОВНАЯ ЧАСТЬ

Данные табл. 1 могут иметь различную физическую природу, поэтому проведем их нормировку. Для этого определим общие средние значения и выборочные дисперсии для каждого из параметров:

$$\overline{y_i} = \frac{1}{n_1 + n_2}\left[\sum_{k=1}^{n_1} y_{ik}^{(1)} + \sum_{k=1}^{n_2} y_{ik}^{(2)}\right] = \frac{n_1}{n_1 + n_2}\overline{y_i^{(1)}} + \frac{n_2}{n_1 + n_2}\overline{y_i^{(2)}},$$

$$D_i^* = \frac{1}{n_1 + n_2}\left[\sum_{k=1}^{n_1} (y_{ik}^{(1)} - \overline{y_i^{(1)}})^2 - \sum_{k=1}^{n_2} (y_{ik}^{(2)} - y_i^{(2)})^2\right],$$

вычислим нормированные данные по формулам

$$x_{ik}^{(1)} = \frac{y_{ik}^{(1)} - \overline{y_i}}{\sqrt{D_i^*}}, \qquad x_{ik}^{(2)} = \frac{y_{ik}^{(2)} - \overline{y_i}}{\sqrt{D_i^*}}$$

Таблица 1

№	Норма				Брак			
	y_1	y_2	...	y_m	y_1	y_2	...	y_m
1	y_{11}	y_{21}	...	y_{m1}	y_{11}	y_{21}	...	y_{m1}
2	y_{12}	y_{22}	...	y_{m2}	y_{12}	y_{22}	...	y_{m2}
...	...	...	...	...	...	...	...	...
n	y_{1n1}	y_{2n1}	...	y_{mn1}	y_{1n2}	y_{2n2}	...	y_{mn2}

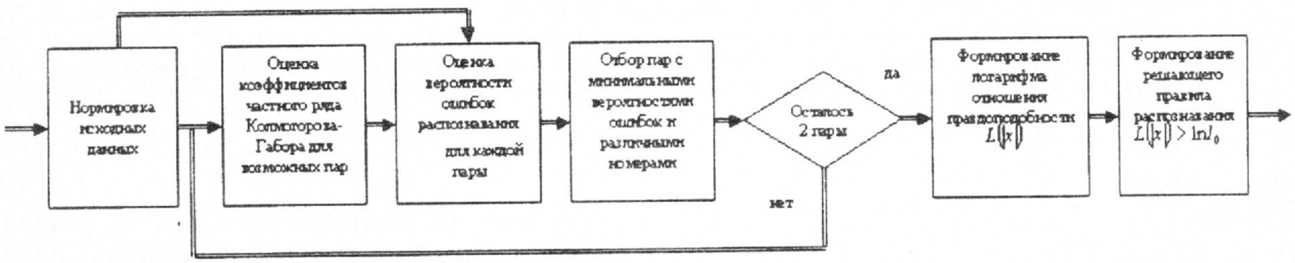

Рис. 1. Алгоритм формирования решающих правил контроля по методу группового учета аргументов

и представим их в виде таблицы, аналогичной табл. 1.

Если неизвестный логарифм отношения правдоподобия

$$L(x_1, x_2, ..., x_m) = \ln\left[\frac{W(x_1, x_2, ..., x_m/H)}{W(x_1, x_2, ..., x_m/Б)}\right]$$

разложить в ряд Тейлора и ограничить его числом параметров m, то получим ряд Колмогорова — Габора

$$L(|x|) = a_0 + \sum_{i=1}^{m} a_i x_i + \sum_{i=1}^{m}\sum_{j=1}^{m} a_{ij} x_i x_j + \sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{m} \dots .$$

Оценив неизвестные коэффициенты этого ряда, получим эмпирическое правило распознавания:

1) если выполняется неравенство

$$L(|x|) - \ln l_0 > 0$$

то контролируемый объект относится к норме;

2) если выполняется противоположное неравенство

$$L(|x|) - \ln l_0 < 0,$$

то объект контроля относиться к браку.

Для определения коэффициентов ряда Колмогорова — Габора составим уравнения

$$L(x_{1i}^{(1)}, x_{2i}^{(1)}, ..., x_{mi}^{(1)}) - \ln l_0 = +1, \quad i = 1, 2, ..., n_1, \tag{1}$$

$$L(x_{1j}^{(2)}, x_{2j}^{(2)}, ..., x_{mj}^{(2)}) - \ln l_0 = -1, \quad j = 1, 2, ..., n_2$$

и решим их относительно неизвестных коэффициентов a_0, a_1, a_2, ..., a_m. Однако на практике для решения таких уравнений необходимо иметь много измерений, так как число членов ряда Колмогорова — Габора резко возрастает с увеличением числа измеряемых параметров. Например, если распознавание ведется по двум параметрам, то число членов ряда равно 7, по трем — 20, по четырем — 70, по пяти параметрам — 252 и т. д. Эту трудность можно преодолеть, если воспользоваться методом группового учета аргументов, который был предложен А. Г. Ивахненко [1—3] для моделирования сложных систем и решения задач прогнозирования. Используя идею этого метода, будем последовательно формировать и группировать решающие правила на основе частных полиномов Колмогорова — Габора второго порядка:

$$L(x_i x_j) = a_0 + a_1 x_i + a_2 x_j + a_3 x_i^2 + a_4 x_j^2 + a_5^{x_i} x_j. \tag{2}$$

В задачах распознавания эффективность решающих правил зависит от различий законов распределения параметров, характеризующих нормальное и дефектное состояния объектов контроля. Поэтому можно упростить частные полиномы (2), отбросив квадраты x_1^2 и x_2^2:

$$L(x_i x_j) = a_0 + a_1 x_i + a_2 x_j + a_3^{x_i} x_j.$$

В результате для каждой пары параметров в уравнениях будет только четыре неизвестных коэффициента, которые легко определяются путем решения переопределенной системы уравнений (1).

Алгоритм формирования эмпирического решающего правила по методу группового учета аргументов представлен на рис. 1.

После нормировки исходных данных частные полиномы оцениваются для всех возможных пар параметров, и определяется их эффективность. В качестве показателя эффективности выбраны оценки вероятностей ошибок распознавания. Отбираются наиболее эффективные пары, в которых номера не повторяются. Эти информативные параметры снова группируются по парам, и решение задачи повторяется. Таким образом, отбираются наиболее эффективные параметры по критерию минимума оценок вероятностей ошибок распознавания, затем группируются и используются для построения рабочего решающего правила. Возможная схема отбора и группировки информативных параметров представлена на рис. 2.

Эффективность и работоспособность полученных таким методом эмпирических решающих правил проверялись путем проведения вычислительного эксперимента, блок-схема которого представлена на рис. 3.

В ходе исследований было изучено влияние длины обучающих выборок на эффективность распознавания. Результаты приведены в табл. 2, где $P_{opt}11$ — оценка вероятности распознавания нормы по оптимальному решающему правилу при условии, что объект контроля находится в норме; $P_{opt}22$ — оценка вероятности распознавания брака по оптимальному решающему правилу при условии, что объект контроля относится к браку; $P_{\text{МГУА}}11$ и $P_{\text{МГУА}}22$ — оценки вероятности распознавания нормы и брака по эмпирическому решающему правилу на основе метода группового учета аргументов.

Из анализа данных табл. 2 следует, что длина исходных выборок влияет на эффективность распознавания: чем длиннее выборки, тем больше значения оценок вероятностей распознавания и тем ближе они к оптимальным значениям. Однако также важно отметить, что метод работает и в условиях ограниченности исходных данных.

На рис. 4 показано изменение средней вероятности распознавания в зависимости от длины обуча-

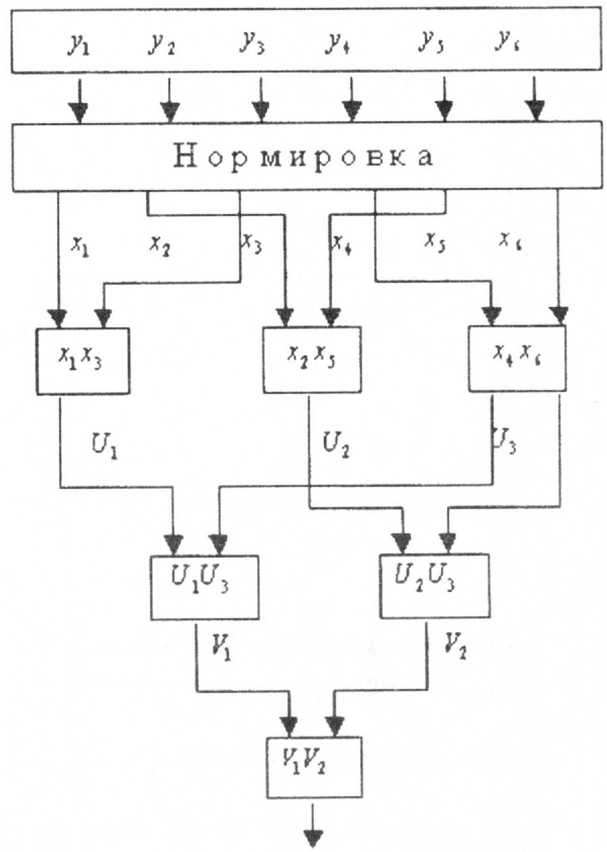

Рис. 2. Возможная схема отбора и группировки информативных параметров

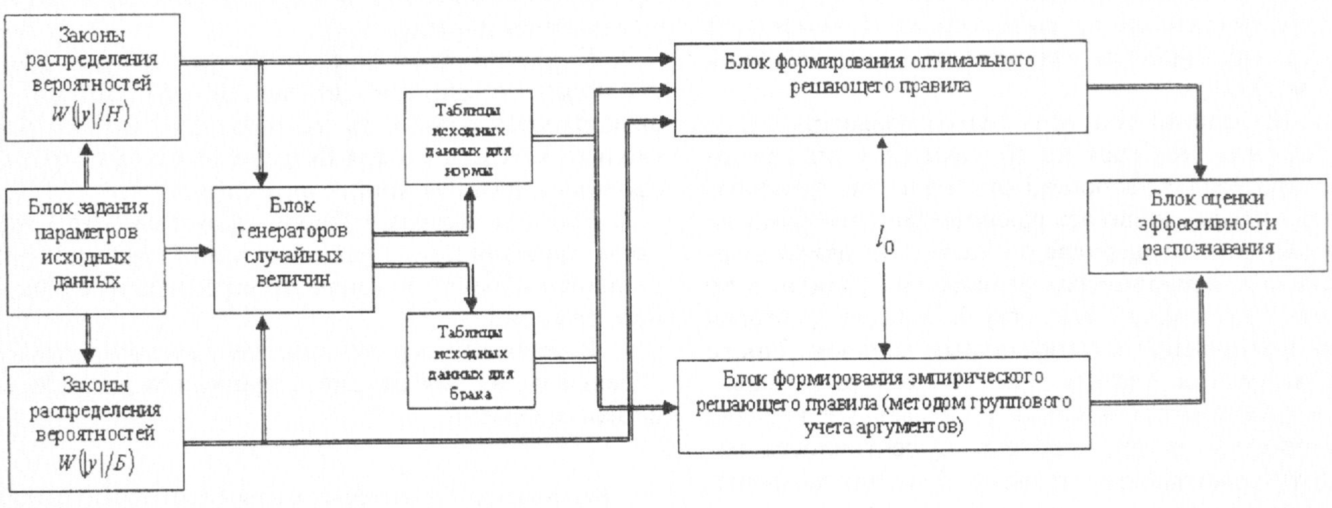

Рис. 3. Блок-схема вычислительного эксперимента

Таблица 2

Длина выборок, $n_{o1} = n_{o2}$	$P_{opt}11$	$P_{opt}22$	$P_{МГУА}11$	$P_{МГУА}22$
10	0.924	0.932	0.914	0.884
25			0.923	0.886
50			0.911	0.901
100			0.913	0.905

Таблица 3

Длина обуч. выборок	Длина провер. выборок	$P_{opt}11$	$P_{opt}22$	$P_{МГУА}11$	$P_{МГУА}22$
		Длинные выборки			
90	10	0.88	0.98	0.87	0.96
70	30	0.89	0.93	0.91	0.9
50	50	0.91	0.94	0.92	0.9
30	70	0.93	0.94	0.92	0.91
10	90	0.91	0.94	0.87	0.91
{100}	{100}	{0.94}	{0.95}	{0.93}	{0.91}
		Короткие выборки			
5	20	0.9	0.85	0.84	0.6
10	15	0.86	0.87	0.8	0.8
15	10	0.91	0.9	0.89	0.81
20	5	0.9	0.8	0.87	0.8
{25}	{25}	{0.92}	{0.92}	{0.91}	{0.92}

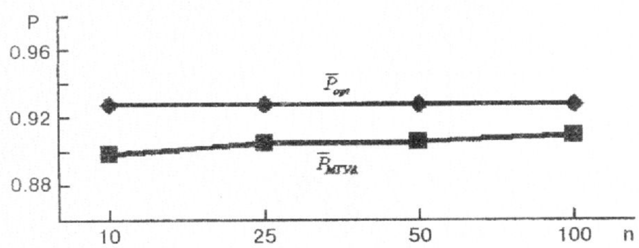

Рис. 4. График зависимости средней вероятности распознавания от длины выборки

ющей выборки. На графике видно, что эффективность распознавания по выборкам из 10 измерений меньше оптимальной вероятности распознавания всего на 3 %.

Был исследован вопрос о целесообразности деления исходных выборок на обучающие и проверяющие. Проведен следующий эксперимент: фиксированная выборка делилась различными способами на обучающую и проверяющую части. По одной формировалось эмпирическое решающее правило, а по другой проверялась его эффективность, которая затем сравнивалась с оптимальным каналом. Также были получены оценки вероятностей распознавания в случае использования всей выборки как для обучения, так и для проверки эффективности построенного решающего правила. Результаты вычислительных экспериментов, проведенных для длинных и коротких выборок, представлены в табл. 3.

Из анализа полученных данных следует вывод, что наилучшее распознавание будет в случае, если пользоваться всей выборкой и для обучения, и для проверки. Это особенно важно при формировании решающих правил по коротким выборкам.

ВЫВОДЫ

1. Использование метода группового учета аргументов позволяет решать задачу отбора информативных параметров и формирования решающих правил распознавания непосредственно по экспериментальным данным.
2. К достоинствам данного метода следует отнести простоту алгоритма вычислений: от ряда к ряду выполняются одни и те же операции столько раз, сколько необходимо для формирования решающего правила в каждом конкретном случае.
3. Длина исходных выборок влияет на эффективность решающего правила распознавания: чем длиннее выборки, тем меньше вероятности ошибочных решений.
4. В условиях ограниченности исходных данных деление их на обучающие и проверяющие выборки нецелесообразно.

1. Ивахненко А. Г. Самообучающиеся системы распознавания и автоматического регулирования. — К.: Техника, 1969.— 392 с.

2. Ивахненко А. Г. Метод группового учета аргументов в задачах прогнозирования. — К.: Ин-т кибернетики АН УССР, 1976.—25 с.
3. Ивахненко А. Г., Мюллер Й. А. Самоорганизация прогнозирующих моделей. — К.: Техника, 1985.—223 с.
4. Малайчук В. П., Мозговой А. В. Обработка информации в средствах и системах неразрушающего контроля. — Д.: Изд-во ДГУ, 1992.—168 с.
5. Малайчук В. П., Мозговой А. В., Петренко А. Н. Математический аппарат решения задач неразрушающего контроля // Матер. десятой юбилейной междунар. конф. и выставки «Современные методы и средства неразрушающего контроля и технической диагностики». — К.: УИЦ «Наука. Техника. Технология», 2002.—С. 76—87.

THE DERIVATION OF DECISION RULES OF NON-DESTRUCTIVE TESTING ON THE BASIS OF THE METHOD OF THE GROUP REGISTRATION OF ARGUMENTS

N. Lysenko

The efficiency of the empirical decision rules constructed on the basis of the method of the group registration of arguments is investigated. The comparison of the empirical and optimum decision rules is carried out by computing experiments. The estimates for the probabilities of the identification are derived and their relation to the length of the initial sampling is investigated. The problem of the division of the sampling into learning and checking ones is analysed.

УДК 004.056(075.8)

© **Е. В. Дзыгин**

Дніпропетровський національний університет

АДАПТИВНЫЙ АЛГОРИТМ СЖАТИЯ НЕСТАЦИОНАРНЫХ ЗНАКОВЫХ СООБЩЕНИЙ

Розроблено адаптивний алгоритм стиснення нестаціонарних знакових повідомлень. Експериментально досліджено зміну коефіцієнта стиснення у залежності від довжини і типу повідомлення. Розроблено адаптивний алгоритм стиснення. Отримано практичні рекомендації з виявлення змін статистики стискуваного повідомлення. Проведено порівняння розробленого алгоритму стиску з базовими алгоритмами.

В настоящее время есть множество алгоритмов сжатия, которые хорошо зарекомендовали себя при сжатии стационарных знаковых сообщений. В случае, если статистика сообщения нестационарна, то степень сжатия значительно уменьшается. Эта проблема актуальна ввиду появившейся в современных компьютерных системах тенденции к комбинированию в одном файле различных типов данных (тексты на различных языках, аудиоданные, графическая информация и т. п.), служащих для решения одной задачи. Здесь предлагается адаптивный алгоритм сжатия, определяющий изменения в статистике сообщения с помощью отслеживания изменения коэффициента сжатия.

В качестве базовых алгоритмов используется адаптивный алгоритм Хаффмана и алгоритм LZW. Алгоритм Хаффмана является статистическим, а LZW — словарным. В статистическом сжатии каждому символу присваивается код, основанный на вероятности его появления в тексте. Высоковероятные символы получают короткие коды, и наоборот. В словарном методе группы последовательных символов или «фраз» заменяются кодом. Замененная фраза может быть найдена в некотором «словаре».

Сжатие может осуществляться по априорным данным о статистических свойствах исходных сообщений, или же проводиться по технологии индивидуального кодирования каждого сообщения. Адаптивный алгоритм Хаффмана и алгоритм LZW реализуют индивидуальное кодирование для каждого сообщения. В процессе своей работы они приспосабливаются к статистике сообщения, и качество сжатия улучшается.

Идея написания адаптивного алгоритма сжатия возникла после проведения ряда исследований изменения коэффициента сжатия от длины прочитанного сообщения. Коэффициент сжатия определяется как отношение длины исходного сообщения к длине сжатого сообщения:

$$K = Nl/L,$$

где N — количество знаков в исходном сообщении, l — длина одного знака сообщения, L — длина сжатого сообщения.

На рис. 1 представлен график изменения коэффициента сжатия от длины прочитанного сообщения. Для сжатия применялся текст на английском языке. Длина сообщения 1700 знаков. Видно, что на начальном этапе сжатия использовать алгоритм LZW невыгодно, поскольку приблизительно до 500