DOI 10.17587/issn.1684-6400

ISSN 1684-6400

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

Том 26

2020

№ 10

ТЕОРЕТИЧЕСКИЙ И ПРИКЛАДНОЙ НАУЧНО-ТЕХНИЧЕСКИЙ ЖУРНАЛ

САПР

КОМПЬЮТЕРНАЯ ГРАФИКА

МЕТОДЫ ПРОГРАММИРОВАНИЯ

ОПЕРАЦИОННЫЕ СИСТЕМЫ И СРЕДЫ

ТЕЛЕКОММУНИКАЦИИ И ВЫЧИСЛИТЕЛЬНЫЕ СЕТИ

ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ

НЕЙРОСЕТИ И НЕЙРОКОМПЬЮТЕРЫ

СТРУКТУРНЫЙ СИНТЕЗ

ВЫЧИСЛИТЕЛЬНЫЕ СИСТЕМЫ

ПРИКЛАДНЫЕ ИНФОРМАЦИОННЫЕ СИСТЕМЫ

ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ

ОПТИМИЗАЦИЯ И МОДЕЛИРОВАНИЕ

ИТ В ОБРАЗОВАНИИ

ГИС

Рисунки к статье А. Ю. Спасёнова, К. В. Кучерова, Т. М. Волосатовой, Д. М. Жука

«ОЦЕНКА СОСТОЯНИЯ СЛОЖНЫХ ТЕХНИЧЕСКИХ ОБЪЕКТОВ С ИСПОЛЬЗОВАНИЕМ СТРУКТУРНО-МОДАЛЬНОГО АНАЛИЗА КВАЗИПЕРИОДИЧЕСКИХ ВРЕМЕННЫХ РЯДОВ»

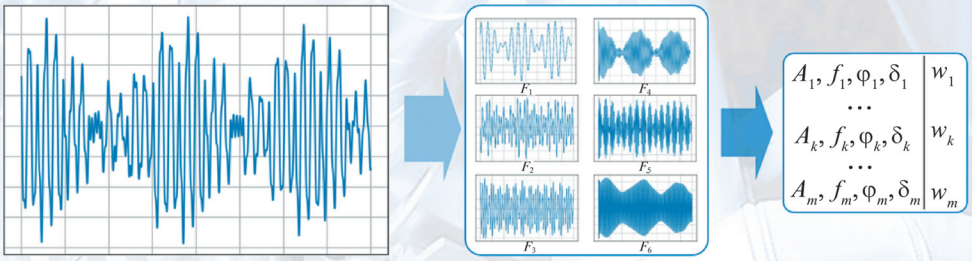


Рис. 3. Модальная декомпозиция сигнала

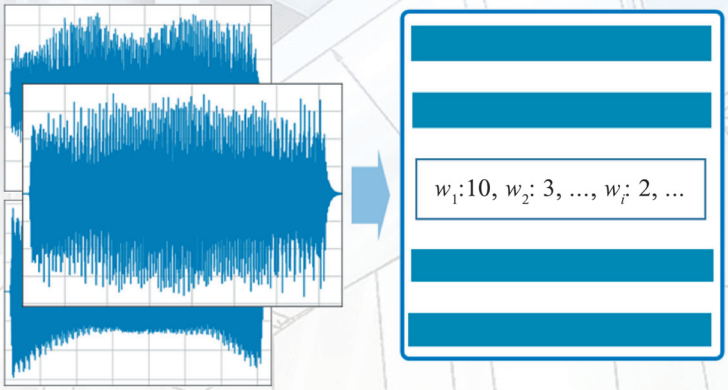


Рис. 4. Портрет состояния системы

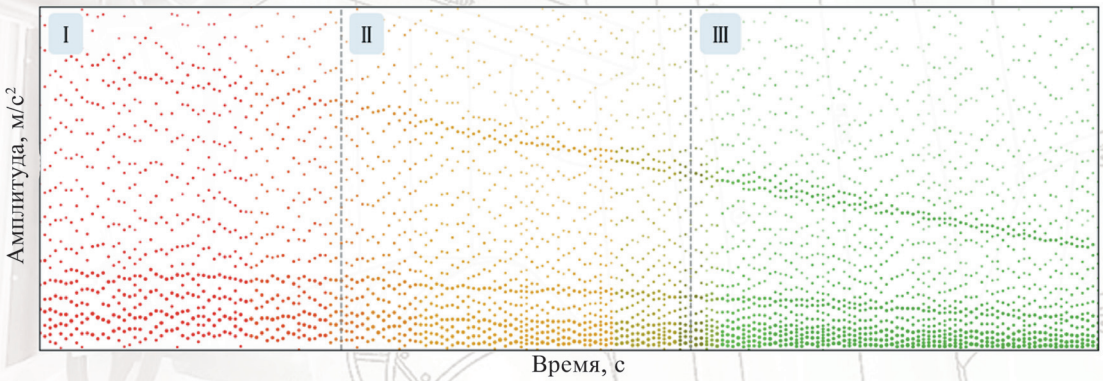


Рис. 5. Изменение амплитуд сегментов

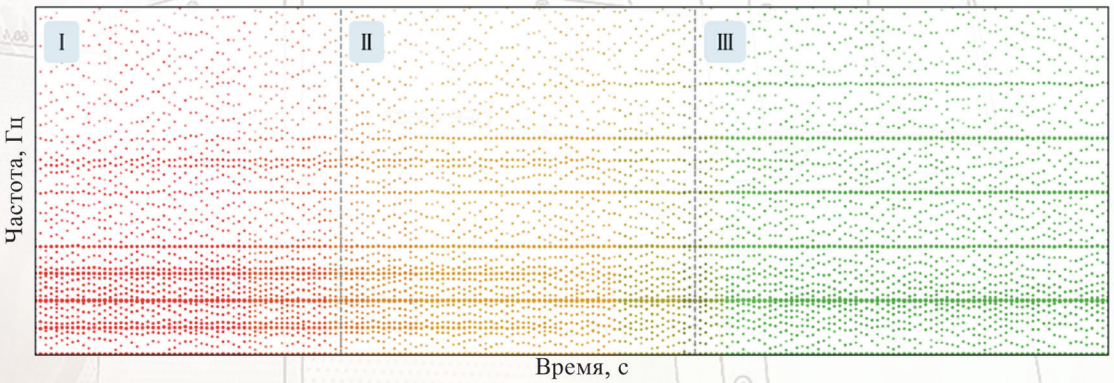


Рис. 6. Изменение частот сегментов

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

Том 26
2020
№ 10

ТЕОРЕТИЧЕСКИЙ И ПРИКЛАДНОЙ НАУЧНО-ТЕХНИЧЕСКИЙ ЖУРНАЛ

Издается с ноября 1995 г. DOI 10.17587/issn.1684-6400

УЧРЕДИТЕЛЬ
Издательство "Новые технологии"

СОДЕРЖАНИЕ

Информация о журнале доступна по сети Internet по адресу **http://novtex.ru/IT.**
Журнал включен в систему Российского индекса научного цитирования и базу данных RSCI на платформе Web of Science.
Журнал входит в Перечень научных журналов, в которых по рекомендации ВАК РФ должны быть опубликованы научные результаты диссертаций на соискание ученой степени доктора и кандидата наук.

INFORMATION TECHNOLOGIES
INFORMACIONNYE TEHNOLOGII

Vol. 26
2020
No. 10

THEORETICAL AND APPLIED SCIENTIFIC AND TECHNICAL JOURNAL

Published since November 1995

DOI 10.17587/issn.1684-6400

ISSN 1684-6400

CONTENTS

Complete Internet version of the journal at site: **http://novtex.ru/IT.**
According to the decision of the Higher Certifying Commission of the Ministry of Education of Russian Federation, the journal is inscribed in "The List of the Leading Scientific Journals and Editions wherein Main Scientific Results of Theses for Doctor's or Candidate's Degrees Should Be Published"

УДК 519.25 DOI: 10.17587/it.26.555-563

В. Г. Зайцев, вед. инженер, e-mail: zaycev@looch.ru,
Научно-производственное предприятие геофизической аппаратуры "Луч", г. Новосибирск

К оцениванию параметров моделей при наличии ошибок во входной и выходной переменных. Линейный случай

Рассматриваются способы идентификации систем, описываемых моделями с погрешностями регистрации входной и выходной переменных. Считается, что при отсутствии информации о параметрах погрешностей в случае больших их значений получить приемлемые оценки всех искомых параметров модели без дополнительных, в сравнении с регрессионным анализом, предположений нельзя. Определение параметров регрессии при наличии погрешностей во входных и выходных переменных актуально при решении многих задач обработки данных. Во всех существующих подходах либо принимаются допущения относительно помех в виде, например, заданных соотношений между дисперсиями помех, либо используется метод последовательных приближений.

Для определения параметров линейной зависимости предложено использовать условие симметрии совместной плотности вероятности наблюдаемых входной и выходной переменных в косоугольных координатах.

Показана состоятельность и несмещенность оценки параметра связи переменных. Для случая нормальности помех в статье приведена формула определения параметра связи через оценки семиинвариантов четвертого порядка.

Ключевые слова и фразы: *структурный анализ, симметричное распределение помех, косоугольная система координат, состоятельность оценки*

Введение

Рассматриваются способы идентификации систем, описываемых моделями с погрешностями регистрации входной и выходной переменных. Если обе переменные системы случайны, то определение связей между ними обычно называют структурным анализом, а если входная переменная неслучайна — конфлюентным анализом. Считается, что при отсутствии информации о параметрах погрешностей в случае больших их значений получить приемлемые оценки всех искомых параметров модели без дополнительных, в сравнении с регрессионным анализом, предположений нельзя [1, 2]. Во всех существующих подходах либо принимаются допущения относительно помех в виде, например, заданных соотношений между дисперсиями помех, либо используется метод последовательных приближений.

Идентификация в одномерном линейном случае

Рассмотрим подробно возможные методы идентификации для одномерного случая. Пусть случайные величины (СВ) W и V связаны однозначной зависимостью $W = \Psi(V)$ и в пассивном эксперименте вместо W и V наблюдаются СВ

$$Y = W + E, \quad (1)$$
$$X = V + H, \quad (2)$$

где E и H — случайные независимые нормально распределенные ошибки измерений со свойствами $M[E] = 0$, $M[H] = 0$; $M[HE] = 0$, $M[HV] = 0$, $M[EV] = 0$, где M — символ математического ожидания.

Обозначим дисперсии СВ E, H и V соответственно $M[E]^2 = \sigma_\varepsilon^2$; $M[H]^2 = \sigma_\eta^2$; $M[V]^2 = \sigma_v^2$, а математическое ожидание СВ V обозначим $M[V] = m_v$.

Пусть в линейном случае структура системы задана соотношением

$$W = a + \theta V, \quad (3)$$

где a и θ — неизвестные коэффициенты, их значения необходимо оценить по выборке СВ X и Y объемом n. По наблюдаемым данным можно найти оценки параметров $R_{xy} = \theta\sigma_v^2$; $\sigma_x^2 = \sigma_v^2 + \sigma_\eta^2$; $\sigma_y^2 = \theta\sigma_v^2 + \sigma_\varepsilon^2$; $m_x = m_v$, $m_y = a + \theta m_v$, где R_{xy} — коэффициент ковариации; m_x, m_y — средние значения СВ X и Y. Модель не идентифицируется, поскольку для шести не-

известных a, θ, $\sigma_\varepsilon^2, \sigma_\eta^2, \sigma_v^2$ и m_v имеем пять уравнений. Вычтем из СВ X и Y их средние значения и преобразуем СВ. В геометрической интерпретации преобразование представляет собой повороты осей Ox, Oy исходной системы координат (СК) к новым осям Ox_{01}, Oy_0:

$$X_{01} = [(X - m_x)\cos\beta + (Y - m_y)\sin\beta]\cos^{-1}(\alpha - \beta); \quad (4)$$

$$Y_0 = [-(X - m_x)\sin\alpha + (Y - m_y)\cos\alpha]\cos^{-1}(\alpha - \beta), \quad (5)$$

где α и β — углы поворота осей, причем $\mathrm{tg}\alpha = \theta$, а $\mathrm{tg}\beta = \mathrm{tg}\alpha\sigma_\eta^2\sigma_\varepsilon^{-2}$. Повороты оси Ox к оси Ox_0 на угол α и оси Oy к оси Oy_0 на угол β выполняются против хода движения часовой стрелки.

Подставив в соотношения (4) и (5) выражения (1)—(3), получим

$$X_{01} = [V - m_x + (a + \theta V - m_y)\mathrm{tg}\beta]\cos^{-1}\alpha\cos^{-1}(\alpha - \beta) + H_0;$$

$$Y_0 = [-(V - m_x)\mathrm{tg}\alpha + (a + \theta V - m_y)]\cos^{-1}\beta\cos^{-1}(\alpha - \beta) + E_0,$$

где $H_0 = (H\cos\beta + E\sin\beta)\cos^{-1}(\alpha - \beta)$ и $E_0 = (-H\sin\alpha + E\cos\alpha)\cos^{-1}(\alpha - \beta)$.

Проведя тригонометрические преобразования, с учетом равенства $\mathrm{tg}\alpha = \theta$ придем к выражениям

$$X_{01} = H_0 + V_0 + c_1;$$

$$Y_0 = E_0,$$

где $V_0 = V\cos^{-1}\alpha$, $c_1 = [(a - m_y)\mathrm{tg}\beta - m_x]\cos^{-1}\alpha \times \cos^{-1}(\alpha - \beta)$.

Перенесем параллельно себе ось Ox_{01} так, чтобы $X_0 = X_{01} - c_1$. Математическое ожидание произведения равно $M[H_0E_0] = M[(H\cos\beta + E\sin\beta)(-H\sin\alpha + E\cos\alpha)]\cos^{-2}(\alpha - \beta) = 0$. Сумма нормальных величин — величина нормальная, следовательно, является СВ. H_0 и E_0 независимы.

Рассмотрим плотность вероятности (ПВ) $f_{X_0Y_0}(x, y)$ системы СВ (X_0, Y_0). Эту плотность считаем симметричной относительно осей Ox_0, Oy_0, если

$$f_{X_0Y_0}(x, 0) = f_{X_0Y_0}(-x, 0) \quad (6)$$

$$\text{и} \quad f_{X_0Y_0}(0, y) = f_{X_0Y_0}(0, -y). \quad (7)$$

В связи с независимостью СВ H_0 и E_0 ПВ $f_{H_0E_0}(\eta, \varepsilon)$ системы (H_0, E_0) будет симметрична по аналогии с соотношениями (6), (7).

Теорема 1. *Если плотность СВ V симметрична относительно математического ожидания, то при нормальной ПВ $f_{HE}(\eta, \varepsilon)$ существует единственная система координат с осями Ox_0, Oy_0, в которой ПВ $f_{X_0Y_0}(x, y)$ системы СВ (X_0,Y_0) симметрична относительно осей Ox_0 и Oy_0, одна из которых совпадает с линией $y = \theta x$. Если ПВ $f_V(v)$ несимметрична, то лишь в единственной указанной системе координат ПВ системы (X_0, Y_0) симметрична относительно оси, проходящей через линию $y = \theta x$.*

Во многих практических задачах регрессионного анализа при отсутствии погрешности независимой переменной условия применимости метода наименьших квадратов (МНК) нарушаются. Распространенным нарушением этих условий является неоднородность дисперсий наблюдаемой величины Y — гетероскедастичность данных. Встречаются ситуации, когда данные неоднородны по дисперсии, но их можно разделить на несколько групп однородных. Гетероскедастичность преодолевается с использованием взвешенного МНК (ВМНК) с многоэтапной процедурой оценивания коэффициентов регрессии на основе обычного МНК.

Несмотря на множество примеров использования ВМНК помеха E зависимой переменной математически в них не описывается: если она имеет различную дисперсию для различных значений независимой переменной, то формально она не случайная величина, а вектор. Однако это не мешает использовать ВМНК без формального определения этой помехи как вектора. Рассмотрим m неоднородных групп наблюдений, но с погрешностями в независимой и зависимой переменных.

Предположим, что СВ V в соотношениях (2) и (3) является дискретной величиной v_i, $i = 1, ..., m$, неважно, известной или нет. Число m этих значений может быть неизвестным. Свяжем с каждой парой v_i и $w_i = \theta v_i$ плотность вероятности $f_{iHE}(\eta, \varepsilon) = f_{iH}(\eta)f_{iE}(\varepsilon)$, при этом дисперсии нормально распределенных помех $\sigma_{i\eta}^2$ и $\sigma_{i\varepsilon}^2$ будем считать неизвестными, а их отношения $\sigma_{i\eta}^2/\sigma_{i\varepsilon}^2$ также неизвестным, но одинаковыми для всех i. Предельный случай подобного варианта: наличие m известных дискретных значений $v_i = i\Delta v$, где Δv — интервал между наблюдениями; $\sigma_{i\eta}^2 = 0$; $\sigma_{i\varepsilon}^2 = \sigma_\varepsilon^2 = \mathrm{const}$, $i = 1, ..., m$. Это представление наиболее часто используется в практике регрессионного анализа.

Следствие 1. *Теорема 1 выполняется в случае, если V представляет собой набор дискретных неизвестных значений v_i, $i = 1,...,m$, каждому из которых соответствует плотность ошибок $f_{iHE}(\eta,\varepsilon)$ с неизвестными в каждой точке значениями дисперсий $\sigma_{i\eta}^2$ и $\sigma_{i\varepsilon}^2$, но с одинаковым неизвестным отношением $\sigma_{i\eta}^2/\sigma_{i\varepsilon}^2$.*

Возможны различные алгоритмы получения оценок параметра θ с использованием

свойств симметрии ПВ $f_{XY}(x, y)$. Эту оценку можно определить, сравнивая непосредственно эмпирическую плотность $f_{eXY}(x, y)$ по разную сторону осей СК.

Теорема 2. *Оценка параметра* θ, *определяемая по минимуму критерия Пирсона как критерия однородности выборок, относящихся к различным квадрантам координатной плоскости, в единственной системе координат, в которой ПВ* $f_{XY}(x, y)$ *симметрична, является состоятельной и асимптотически несмещенной.*

Доказательство теоремы 1. Доказательство проведем вначале для симметричной ПВ $f_v(v)$. До преобразования координат ПВ системы СВ (X, Y) при линейной связи W и V имеем

$$f_{XY}(x, y) = \int_{-\infty}^{\infty} f_V(v) f_H(x - v) f_E(y - \theta v) dv.$$

Поскольку в системе координат с осями Ox_0, Oy_0 СВ H_0 и E_0 также независимы, ПВ системы СВ (X_0, Y_0) равна

$$f_{X_0Y_0}(x_0, y_0) = f_E(y_0) \int_{-\infty}^{\infty} f_V(v_0) f_H(x_0 - v_0) dv_0, \quad (8)$$

где $f_E(\varepsilon_0)$, $f_V(v_0)$, $f_H(\eta_0)$ находятся по правилам определения ПВ функций СВ.

ПВ $f_E(\varepsilon_0)$, $f_H(\eta_0)$ симметричны. Если ПВ $f_v(v_0)$ симметрична, то свертка (8) даст ПВ $f_{X_0Y_0}(x_0, y_0)$, зеркально симметричную относительно осей Ox_0, Oy_0.

Докажем теперь единственность угла $\alpha = \operatorname{arctg}\theta$, определяющего систему координат, в которой ПВ $f_{XY}(x, y)$ симметрична. Рассмотрим произвольную прямую $y = \theta_1 x$, $\theta_1 \neq \theta$ в системе координат с осями Ox_1, Oy_1. Преобразуем СВ X и Y по формулам (4), (5), только при этом вместо $\operatorname{tg}\alpha$ имеем $\operatorname{tg}\alpha_1 = \theta_1$, вместо $\operatorname{tg}\beta$ используем $\operatorname{tg}\beta_1$, а $V_1 = V/\cos\alpha_1$.

По аналогии с получением X_{01}, Y_0, H_0, E_0 преобразуем СВ X и Y:

$$X_{11} = V_1 c_2 + c_3 + H_1; \quad (9)$$

$$Y_{11} = V(\theta - \theta_1)\cos^{-1}(\alpha_1 - \beta_1) + m_x(\theta_1 - \theta)\cos^{-1}(\alpha_1 - \beta_1) + E_1, \quad (10)$$

где

$$c_2 = (\sigma_\varepsilon^2 + \theta\theta_1\sigma_\eta^2)/(\sigma_\varepsilon^2 + \theta_1^2\sigma_\eta^2);$$

$$c_3 = \{(1 + \theta_1^2)[(a - m_y)\theta_1\sigma_\eta^2 - m_x\sigma_\varepsilon^2]\}/(\sigma_\varepsilon^2 + \theta_1^2\sigma_\eta^2).$$

Вычтем из X_{11} и Y_{11} их средние значения, обозначив разности через X_1 и Y_1.

Поскольку X_1 и Y_1 зависимы, то относительно осей Ox_1, Oy_1 ПВ $f_{X_1Y_1}(x_1, y_1)$ не будет симметричной.

Пусть теперь ПВ $f_v(v)$ несимметрична. Плотность $f_{xy}(x_0, y_0)$ определяется также по формуле (8). Так как $f_\varepsilon(y_0)$ симметрична, то ПВ $f_{xy}(x_0, y_0)$ симметрична относительно оси Ox_0, но несимметрична относительно оси Oy_0. Для угла поворота исходной СК $\alpha_1 \neq \alpha$ можно получить формулы, аналогичные формулам (9), (10) и соответствующие им выводы.

Доказательство следствия 1. В геометрической постановке в СК с осями Ox, Oy размещен набор ПВ $p_{i\eta\varepsilon}(x, y)$, $i = 1, ..., m$, с находящимися на линии $y = \theta x$ значениями математических ожиданий и отличающимися только ими от ПВ $f_{i\eta\varepsilon}(\eta, \varepsilon)$. Перейдем к косоугольной СК. Как и в предыдущем случае, положим $\operatorname{tg}\alpha = \theta$, а $\operatorname{tg}\beta = \operatorname{tg}\alpha\sigma_\eta^2\sigma_\varepsilon^{-2}$. Каждая из ПВ $p_{i\eta\varepsilon}(x, y)$ будет теперь расположена на оси Ox_0, и при этом математическое ожидание $M[E_{0i}] = 0$ и плотность симметрична относительно этой оси в данной СК. Таким образом, и в целом ПВ $f_{xy}(x, y)$, полученная усреднением всех ПВ $p_{i\eta\varepsilon}(x, y)$ в каждой точке косоугольной СК, будет в этой системе симметрична относительно оси Ox_0. Можно показать, что если значения v_i, $i = 1, ..., m$, заданы с одинаковым интервалом, и ПВ $f_{i\eta\varepsilon}(\eta, \varepsilon)$ одинаковы, то ПВ $f_{xy}(x, y)$ будет симметрична относительно обеих осей косоугольной СК.

Доказательство теоремы 2. Рассмотрим два положения осей СК: истинное — с осями Ox_0, Oy_0, тангенсами углов наклона относительно исходных осей Ox, Oy $\operatorname{tg}\alpha = \theta$, $\operatorname{tg}\beta = \operatorname{tg}\alpha\sigma_\eta^2\sigma_\varepsilon^{-2}$ и произвольное положение с отклонением осей Ox_1, Oy_1 от осей Ox_0, Oy_0 на углы $\Delta\alpha = \alpha_1 - \alpha$ и $\Delta\beta = \beta_1 - \beta$. Проведем одинаковое разбиение плоскостей этих СК на площадки, зеркально симметричные относительно осей, с числом площадок в каждом квадранте, равном r^2. Поскольку значения ПВ $f_{x_0y_0}(x_0, y_0)$ относительно осей системы координат с осями Ox_0, Oy_0 теоретически должны зеркально совпадать, то ее оценки в зеркально расположенных площадках можно считать принадлежащими выборкам из одной генеральной совокупности. Оценим однородность этих выборок. Для СК с осями Ox_0, Oy_0 статистика

$$\chi^2_{\lambda 0} = \sum_{i=1}^{r}\sum_{j=1}^{r}\left[\frac{(k_{ij}^{(x_0,y_0)} - k_{ij}^{(x_0,-y_0)})^2}{k_{ij}^{(x_0,y_0)} + k_{ij}^{(x_0,-y_0)}} + \frac{(k_{ij}^{(-x_0,y_0)} - k_{ij}^{(-x_0,-y_0)})^2}{k_{ij}^{(-x_0,y_0)} + k_{ij}^{(-x_0,-y_0)}} + \frac{(k_{ij}^{(y_0,x_0)} - k_{ij}^{(y_0,-x_0)})^2}{k_{ij}^{(y_0,x_0)} + k_{ij}^{(y_0,-x_0)}} + \frac{(k_{ij}^{(-y_0,x_0)} - k_{ij}^{(-y_0,-x_0)})^2}{k_{ij}^{(-y_0,x_0)} + k_{ij}^{(-y_0,-x_0)}}\right] \quad (11)$$

асимптотически подчиняется χ^2-распределению с числом степеней свободы $q = 4r^2 - 1$. Здесь $k_{ij}^{(x_0,y_0)}$, $k_{ij}^{(x_0,-y_0)}$, $k_{ij}^{(-x_0,y_0)}$, $k_{ij}^{(-x_0,-y_0)}$, $k_{ij}^{(y_0,x_0)}$, $k_{ij}^{(y_0,-x_0)}$, $k_{ij}^{(-y_0,x_0)}$, $k_{ij}^{(-y_0,-x_0)}$ — числа попаданий в ij-ю площадку по обе стороны от совпадающей оси двух квадрантов; i, j — номера площадок вдоль осей, начиная от центра координат в сторону увеличения абсолютных значений координат.

Преобразуем первое слагаемое из соотношения (11) с учетом того, что для обоих квадрантов

$$k_{ij}^{(x_0,y_0)} = nsf_{eX_0Y_0}(x_{0ji}, y_{0ji}),$$

$$k_{ij}^{(x_0,-y_0)} = nsf_{eX_0Y_0}(x_{0ji}, -y_{0ji}),$$

где $f_{eX_0Y_0}(x_{0ji}, y_{0ji})$, $f_{eX_0Y_0}(x_{0ji}, -y_{0ji})$ — значения эмпирической плотности для ij-й площадки, s — площадь площадки:

$$\chi^2_{\lambda 01} = \\ = ns\sum_{i=1}^{r}\sum_{j=1}^{r}\frac{[f_{eX_0Y_0}(x_{0ij} - \Delta m_{x_0}, y_{0ij} - \Delta m_{y0}) - f_{eX_0Y_0}(x_{0ij} - \Delta m_{x_0}, y_{0ij} - \Delta m_{y0}) + \\ \to \frac{-f_{eX_0Y_0}(x_{0ij} - \Delta m_{x0}, -y_{0ij} + \Delta m_{y0})]^2}{+f_{eX_0Y_0}(x_{0ij} - \Delta m_{x0}, -y_{0ij} + \Delta m_{y0})}, \quad (12)$$

где $\Delta m_{x0} = (\Delta m_x \cos\beta_e + \Delta m_y \sin\beta_e)\cos^{-1}(\alpha_e - \beta_e)$; $\Delta m_{y0} = (-\Delta m_x \sin\alpha_e + \Delta m_y \cos\alpha_e)\cos^{-1}(\alpha_e - \beta_e)$, Δm_x, Δm_y — ошибки определения средних значений, α_e, β_e — углы поворота осей, соответствующие минимуму суммы значений выражений типа (12) для всех четырех сочетаний пар квадрантов.

Разность оценки плотности в симметрично взятых площадках квадрантов состоит из двух составляющих. Первая связана с ошибками оценивания плотности при $\Delta m_x = 0$ и $\Delta m_y = 0$, а вторая — со смещением нулей аргументов теоретической плотности. Ошибки определения средних значений СВ X и Y, а также X_0 и Y_0 с ростом n стремятся к нулю.

В СК с осями Ox_1, Oy_1 плотности $f_{eX_1Y_1}$ в соседних квадрантах отличаются, т. е. в среднем отличаются и числа попаданий в ij-ю площадку по обе стороны от совпадающей оси в формуле типа (11). При наличии средних значений в каждом слагаемом числителя эта статистика подчиняется нецентральному $\chi^2_{\lambda 1}$-распределению. В этом случае в формулу (12) вместо плотности $f_{eX_0Y_0}$ подставляется $f_{eX_1Y_1}$. Если вместо $f_{eX_1Y_1}$ использовать среднее значение на площадке теоретической плотности $f_{X_1Y_1}$, то получаемое при этом неслучайное значение статистики в сумме для четырех сочетаний пар квадрантов будет представлять собой параметр нецентральности $\lambda = 4ns\Delta f$, где Δf — значение суммы (12) при использовании теоретической ПВ $f_{X_1Y_1}$.

Для статистики $\chi^2_{\lambda 0}$ среднее значение и среднеквадратическое отклонение (СКО) равны $m_{\chi 0} = q$ и $\sigma_{\chi 0} = \sqrt{2q}$, а для $\chi^2_{\lambda 1}$ с нецентральным χ^2-распределением — $m_{\chi 1} = q + 4ns\Delta f$, $\sigma_{\chi 1} = \sqrt{2(q + 8ns\Delta f)}$.

Рассмотрим совместную ПВ $f_{\chi_1\chi_0}$ СВ $\chi^2_{\lambda 1}$ и $\chi^2_{\lambda 0}$. Вероятность превышения $\chi^2_{\lambda 0} > \chi^2_{\lambda 1}$ при $\Delta m_x = 0$ и $\Delta m_y = 0$ равна (аргументы x и y соответствуют $\chi^2_{\lambda 1}$ и $\chi^2_{\lambda 0}$ соответственно)

$$P = \int_{y=\chi^2_{\lambda 1}}^{\infty}\int_{x=0}^{\infty} f_{\chi_1\chi_0}(x, y)dxdy. \quad (13)$$

Интегрирование плотности проводится лишь выше линии $y = x$.

Рассмотрим квантили СВ $\chi^2_{\lambda 0}$ и $\chi^2_{\lambda 1}$: $K_{\lambda 0} = m_{\chi 0} + t\sigma_{\chi 0}$ и $K_{\lambda 1} = m_{\chi 1} - t\sigma_{\chi 1}$, где t — постоянная. Найдем их разность

$$\Delta K = 4ns\Delta f - t(\sqrt{2(q + 8ns\Delta f)} + \sqrt{2q}).$$

При $n \to \infty$ $\Delta m_x \to 0$ и $\Delta m_y \to 0$, а также при любом значении t разность $\Delta K \to \infty$. Эта разность характеризует расстояние между ПВ $f_{\chi 1}(x)$ и $f_{\chi 0}(x)$, чем оно больше, тем при меньших значениях пересекаются хвосты этих ПВ и меньше значение интеграла (13), так что при $n \to \infty$ вероятность $P \to 0$.

Таким образом, в случае определения положения осей по минимуму статистики χ^2_λ при $n \to \infty$ отклонения углов от истинных значений $|\Delta\alpha| \to 0$, $\Delta\beta \to 0$, а дисперсия оценки параметра θ также стремится к нулю, что является признаком состоятельности предлагаемой оценки.

Рассмотрим теперь смещенность оценок углов между осями Ox_0, Oy_0 и Ox, Oy, определяемых по минимуму статистики χ^2_λ.

В силу симметричности ПВ $f_{x_0y_0}(x_0, y_0)$ относительно осей, задаваемых углами α и β, совместное распределение ошибок оценивания средних значений Δm_{x0} и Δm_{y0} и ошибок оценки этой ПВ на противоположных относительно осей (оси для несимметричного распределения СВ V) ij-х площадках для бесконечного числа выборок конечного объема одинаковы. Следовательно, распределение всех случайных отклонений $\Delta\alpha$, $\Delta\beta$ будет симметричным относительно истинных значений, оценки углов α и β будут несмещенными. Оценка параметра θ равна $\theta_e = [\theta + \mathrm{tg}(\Delta\alpha)]/[1 - \theta\mathrm{tg}(\Delta\alpha)]$. При $n \to \infty$ $\mathrm{tg}(\Delta\alpha) \to 0$, т. е. оценка θ_e — асимптотически несмещенная.

Примечание. Теорему 1 со следствием и теорему 2 можно распространить и на случай коррелированных ошибок измерений.

Алгоритмы идентификации

Можно предложить несколько алгоритмов получения оценок параметра θ с использованием свойств симметрии ПВ $f_{xy}(x, y)$. Если оценку находить с использованием критерия Пирсона, то общий минимум определяется для суммы минимальных значений, рассчитанных по формуле (11) для всех четырех сочетаний пар квадрантов косоугольной СК, оси которой проведены через средние значения совокупности полученных точек. Оценка параметра a определяется после получения оценки θ_e:

$$a_e = m_{ye} - \theta_e m_{xe}. \qquad (14)$$

При малом объеме выборки использование гистограммы в качестве оценки плотности затруднено. Эмпирическая ПВ СВ X и Y (гистограмма) представляется в виде ступенчатой функции. Если предполагать, что теоретическая ПВ имеет гладкую зависимость от аргументов x и y, то для уменьшения случайных ошибок эмпирической ПВ в виде ступенчатой функции целесообразно применить к ней процедуру сглаживания. При использовании популярного метода сглаживания с помощью ядерных функций возникает проблема выбора параметра локальности ядерной функции в зависимости от объема выборки. Эта проблема решается просто при использовании в качестве ядерных функций непрерывных ПВ без тяжелых хвостов [3]. Рассмотрим СВ

$$R = X + T; \qquad (15)$$

$$Q = Y + U, \qquad (16)$$

где T и U — независимые СВ с нормальными законами распределения, с нулевыми значениями математических ожиданий и заданными значениями дисперсий σ_t^2 и σ_u^2. Обозначим $f_{tu}(t, u)$ совместную плотность распределения СВ T и U. Эту плотность используем в качестве ядерной функции для массива значений выборки

$$f_{exy}(x, y) = n^{-1}\sum_{i=1}^{n} f_{tu}(x - x_i, y - y_i), \qquad (17)$$

где x_i, y_i — значения выборки. Далее с учетом дисперсий σ_t^2 и σ_u^2 уменьшаем масштаб полученной оценки $f_{1xy}(x, y)$, получая сглаженную эмпирическую плотность выборки СВ X и Y с меньшей случайной ошибкой в сравнении со ступенчатой плотностью. Таким образом, за счет информации о гладкости ПВ уменьшается и случайная ошибка определения углов поворота осей.

Определим вид зависимости дисперсии оценки θ_e от характеристик СВ V, E, H и параметра θ. Симметричность ПВ $f_{xy}(x, y)$ в единственной СК означает симметричность в той же СК линий равной плотности вероятности. Рассмотрим произвольное сечение ПВ $f_{xy}(x, y$ и его характерные размеры вдоль осей Ox_0, Oy_0, они пропорциональны СКО СВ X_0 и Y_0. Дисперсии X_0 и Y_0 равны

$$D_{x0} = M[(H\cos\beta + E\sin\beta + V_0)^2] = \\ = (\theta^2 + 1)\sigma_v^2 + \sigma_\eta^2\sigma_\varepsilon^2(\sigma_\varepsilon^2 + \theta^2\sigma_\eta^2)/(\sigma_\varepsilon^4 + \theta^2\sigma_\eta^4), \qquad (18а)$$

$$D_{y0} = M[(-H\sin\alpha + E\cos\alpha)^2] = \\ = (\sigma_\varepsilon^2 + \theta^2\sigma_\eta^2)/(\theta^2 + 1). \qquad (18б)$$

Повернем ось Ox_0 на малые углы $\Delta\alpha$ и $-\Delta\alpha$ и обозначим полученные оси Ox_1 и Ox_2.

Уберем в каждой области горизонтального сечения ПВ $f_{xy}(x, y)$, прилегающей к оси Ox_1, симметричные к этой оси участки. Останутся центральносимметричные участки сечения (участки между осями Ox_1 и Ox_2). Отношение площади этих участков к площади сечения примерно пропорционально отношению СКО σ_{y0}/σ_{x0}. Эти отношения характеризуют рассеяние отклонений углов $\Delta\alpha$ и оценки параметра θ.

При малых углах отклонений найденной прямой $y = \theta_e x$ от истинного угла α дисперсию θ_e можно будет предполагать пропорциональной отношению дисперсий D_{y0}/D_{x0} из соотношений (18а) и (18б).

В качестве одного из используемых условий для первоначального оценивания углов поворота можно принять равенство нулю коэффициента ковариации СВ X_0 и Y_0. Это условие приводит к выражению для определения tgβ:

$$\mathrm{tg}\beta = (D_x\mathrm{tg}\alpha - R_{xy})/(D_y - R_{xy}\mathrm{tg}\alpha). \qquad (19)$$

При этом для некоррелированных СВ E и H полагаем $0 \leqslant \beta \leqslant \pi/2$.

Таким образом, можно подбирать лишь угол α, а угол β будет вычисляться с использованием подбираемого угла α и оценок дисперсий и ковариации. Однако из-за наличия случайности в этих оценках и частного случайных величин можно предполагать несовпадение решений с независимым выбором углов и с использованием формулы (18), а также большее рассеяние результатов для последнего варианта.

Другие способы решения. Наличие информации о моментах погрешностей может упростить решение. Считаем СВ E и H нормальными. Для центрированных СВ $X_c = X - \bar{x}$ и $Y_c = Y - \bar{y}$, где $\bar{x}$ и $\bar{y}$ — математические ожидания, с учетом приведенных выше условий для СВ E и H и их нормальности определим вторые и четвертые моменты:

$$M(X_c^4) = M(V_c^4) + 6M(V_c^4 H^2) + M(H^4) =$$
$$= \mu_{4vc} + 6\sigma_\eta^2\sigma_v^2 + 6K_{v2\eta2} + \mu_{4\eta},$$

где V_c — центрированная входная СВ; μ_{4vc} и $\mu_{4\eta}$ — центральные моменты четвертого порядка СВ V_c и H; $K_{v2\eta2}$ _ обозначение ковариации СВ V_c^2 и H^2. Функции независимых СВ независимы, поэтому $K_{v2\eta2} = 0$. Выражение для четвертого момента примет вид

$$M(X_c^4) = \mu_{4vc} + 6\sigma_\eta^2\sigma_v^2 + 3\sigma_\eta^4.$$

Семиинвариант четвертого порядка СВ X_c равен

$$\rho_4(X_c) = M(X_c^4) - 3\sigma_\chi^4 = \mu_{4vc} - 3\sigma_v^4.$$

Аналогично определяется семиинвариант четвертого порядка СВ Y_c:

$$\rho_4(Y_c) = \theta^4[\mu_{4vc} - 3\sigma_v^4],$$

откуда для оценки параметра имеем

$$\theta_e = \sqrt[4]{\rho_{e4}(Y)/\rho_{e4}(X)}. \qquad (20)$$

В работе [1] модифицированным методом наименьших квадратов для пассивной схемы экспериментов (ММНКП) оценки θ_e определяются с использованием последовательных приближений. На каждом этапе последовательных приближений в ММНКП используется обычный МНК.

Оценки параметров в ММНКП совпадают с оценками, полученными методом наименьших расстояний (МНР) [1, 2]. Утверждается (см., например, работу [2, стр. 244]), что в пассивном эксперименте при неизвестных значениях дисперсий помех построить состоятельные оценки параметра θ нельзя. Оценим в нестрогой постановке ошибки определения оценки θ_e с помощью МНР.

Для одномерной линейной зависимости типа (2) с дискретно заданной величиной v при нахождении оценки коэффициента θ при заданных значениях дисперсий помех в ММНКП с использованием выражения

$$q_e = \operatorname{Arg\,min}\left(n^{-1}\sum_{i=1}^{n}\frac{(y_i - \theta x_i)^2}{\sigma_\varepsilon^2 + \theta^2\sigma_\eta^2}\right) \qquad (21)$$

дисперсию θ_e можно оценить в виде [2]

$$D_{e\theta} = \frac{(\sigma_\varepsilon^2 + \theta^2\sigma_\eta^2)}{\sum_{i=1}^{n}(x_i - m_{xe})^2}. \qquad (22)$$

При некотором объеме выборки бо́льший вклад в ошибку определения параметра вносит не зависящее от n смещение оценки. Считая плотность $f_v(v)$ симметричной, рассмотрим ПВ $f_{xy}(x, y)$ в такой прямоугольной СК с осями Ox_1 и Oy_1, в которой Ox_1 совпадает с прямой $y = \theta x$. ПВ $f_{xy}(x,y)$ получается суммированием бесконечного числа взятых с весом $f_v(v)dv$ плотностей $f_{\eta\varepsilon}(x, y)$ с математическими ожиданиями m_x и $m_y = \theta v$. Если СВ E и H некоррелированы, то при $\sigma_\varepsilon^2 \neq \sigma_\eta^2$ оси симметрии каждого горизонтального сечения ПВ $f_{\eta\varepsilon}(x, y)$ не совпадают с осями Ox_1 и Oy_1. Поэтому оси симметрии горизонтальных сечений плотности $f_{xy}(x, y)$ также не совпадают с этими осями. Уберем в каждой области горизонтального сечения ПВ $f_{xy}(x, y)$, прилегающей к оси Ox_1, симметричные к этой оси участки. Из-за наличия оставшихся после этого по обе стороны от оси центральносимметричных участков сечения можно сделать вывод, что определяемая для каждого сечения прямая $y = \theta_1 x$ с использованием МНР не совпадает с осью Ox_1. Соответственно и для ПВ $f_{x_1y_1}(x, y)$ в данной прямоугольной СК параметр θ_1 не равен нулю, и оценка θ_e по МНР смещена, кроме случаев, когда $\sigma_\varepsilon^2 = \sigma_\eta^2$ или когда одна из осей симметрии горизонтальных сечений ПВ $f_{\eta\varepsilon}(x, y)$ перпендикулярна к линии $y = \theta x$ при коррелированности СВ E и H.

В работе [4] предложен способ, основанный на знании коэффициента эксцесса погрешности η входной переменной. Для определения коэффициента θ этот метод в одномерном случае в предположении нормальности СВ η можно свести к использованию формулы

$$\theta_e = \theta_{11}(\theta_{11}\theta_{12}\mu_{4ex} - 3R_{exy}^2)/(\theta_{11}\mu_{4ex} - 3R_{exy}^2), \quad (23)$$

где μ_{4ex} — оценка четвертого центрального момента СВ X; θ_{11} — МНК-оценка коэффициента θ; θ_{12} — оценка коэффициента θ, полученная взвешенным МНК.

В работе [5] предложена оценка с использованием смешанных моментов:

$$\theta_e = M[YYX]/M[YXX]. \qquad (24)$$

Недостаток двух последних подходов состоит в большей относительно МНР дисперсии оцен-

ки θ_e, в основном обусловленной наличием частного оценок.

В качестве примера рассмотрим последнюю формулу. Распишем ее для конечного объема выборки:

$$\theta_e = \frac{\sum_{i=1}^{n}(\theta v_i + \varepsilon_i)^2(v_i + \eta_i)}{\sum_{i=1}^{n}(\theta v_i + \varepsilon_i)(v_i + \eta_i)^2}.$$

С использованием линеаризации функции случайных величин $Z = \varphi(Z_1, Z_2)$ для среднего значения и дисперсии получим

$$\overline{z} = \overline{z}_1/\overline{z}_2;$$
$$\sigma_z^2 = \sigma_{z2}^2/\overline{z}_2^2 -$$
$$-(\overline{z}_1/\overline{z}_2^2)^2\sigma_{z1}^2 - 2(\overline{z}_1/\overline{z}_2^3)K_{z1z2}.$$

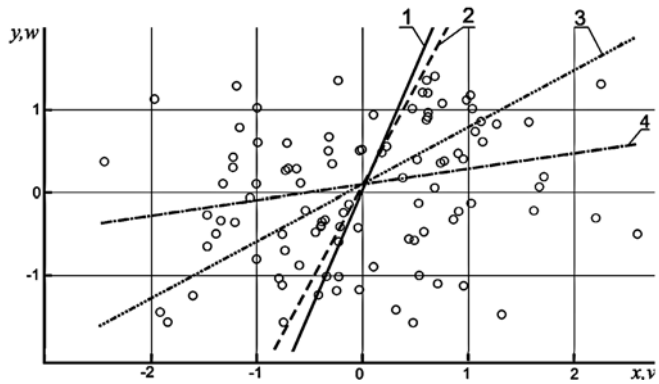
При равном нулю математическом ожидании СВ V оценка θ_e теоретически имеет распределение, для которого среднее значение и дисперсия не существуют. На практике это означает возможность получения больших выбросов этой оценки при малых значениях математических ожиданий СВ V.

Таблица 1

Расчет параметра связи при неслучайной входной величине

	Использование симметрии ПВ		МНР	$\theta_e = \sqrt[4]{\rho_{e4}(Y)/\rho_{e4}(X)}$	$\theta_e = M[YYX]/M[YXX]$	По формуле (23)
	С формулой (19)	Совместный выбор углов				
$\sigma_\varepsilon = 0{,}173$, $\sigma_\eta = 0{,}173$						
$\overline{\theta}$	1,0024	1,0075	1,0087	1,0101	0,1961	0,9670
σ_θ	0,0747	0,0520	0,0513	0,0756	1,3120	0,1328
m_{e2}	0,0056	0,0028	0,0027	0,0058	2,3611	0,0187
$\sigma_\varepsilon = 0$, $\sigma_\eta = 0{,}173$						
$\overline{\theta}$	0,9824	1, 0049	0,9578	1,0004	0,5102	1,0103
σ_θ	0, 0493	0,0285	0,0401	0,0574	2,7890	0,1012
m_{e2}	0,0027	0,0008	0,0034	0,0033	8,0285	0,0102
$\sigma_\varepsilon = 0{,}173$, $\sigma_\eta = 0$						
$\overline{\theta}$	1,0226	1,0077	1,0444	1,0139	1,5850	0,9980
σ_θ	0,0480	0,0291	0,0342	0,0464	1,1511	0,0935
m_{e2}	0,0028	0,0009	0,0031	0,0024	1,6641	0,0087
$\sigma_\varepsilon = 0$, $\sigma_\eta = 0{,}5$						
$\overline{\theta}$	0,9531	0,9902	0,7146	1,0313	0,6911	1,0364
σ_θ	0,1323	0,0911	0,0745	0,3135	2,6338	2,7354
m_{e2}	0,0197	0,0083	0,0870	0,0993	7,0326	7,4839

Примеры. 1. Входную величину v брали с интервалом 0,04 в диапазоне от $-1{,}0$ до 1,0. Были промоделированы нормально распределенные помехи с объемом выборки $n = 51$. Дисперсии погрешностей принимали равными $\sigma_\varepsilon^2 = 0{,}03$ и $\sigma_\eta^2 = 0{,}03$; $\sigma_\varepsilon^2 = 0$ и $\sigma_\eta^2 = 0{,}03$; $\sigma_\varepsilon^2 = 0{,}03$ и $\sigma_\eta^2 = 0$; $\sigma_\varepsilon^2 = 0$ и $\sigma_\eta^2 = 0{,}25$. Искомые параметры принимали равными $a = 0$ и $\theta = 1$. В табл. 1 представлены результаты расчетов методом ММНКП/МНР, методом оценивания параметра θ, основанном на теоретическом отношении $\theta = M(YYX)/M(YXX)$, методом из работы [4] по формуле (23) и предложенными методами. При этом оси под углами α и β относительно старых осей проходят через точку, соответствующую средним значениям выборки СВ X и Y. B каждую точку (x_0, y_0) косоугольной системы координат, получаемую перебором углов поворота обеих осей, помещалась двумерная плотность нормального распределения с дисперсией по каждой координате $\sigma^2 = (\sigma_x^2 + \sigma_y^2)n^{0,66}$ и числом интервалов по каждой координате, равным 40. Для нахождения минимального значения критерия Пирсона в каждом положении осей суммировались квадраты разностей сглаженной ядерным способом оценки $f_{exy}(x_0, y_0)$ в четырех парах соседних квадрантов. Число значений выборок было равно 100. Для различных значений СКО ошибок в табл. 1 и в табл. 2 помещены среднее значение $\overline{\theta}$, СКО σ_θ и второй центральный момент m_{e2} оценки. Если СКО ошибок принимались одинаковыми, предлагаемый метод с перебором углов поворота обеих осей давал результа-



Истинная прямая (*1*); прямая, полученная предлагаемым методом (*2*); прямая, полученная МНР (*3*); прямая, полученная МНК (*4*)

ты, близкие к получаемым МНР. Однако при $\sigma_\varepsilon^2 \neq \sigma_\eta^2$ МНР ему существенно уступает.

Для МНР вначале использовали алгоритм из работы [2], а после этого, для увеличения точности оценки, в окрестности найденного значения проводили перебор значений θ с поиском минимальной суммы абсолютных значений отклонений исходных точек от прямой линии $y = a + \theta x$.

Средние значения оценки параметра a по формуле (14) с использованием предложенных в статье способов, а также МНР дали значения 0,001...0,004, а СКО оценок 0,02...0,08.

2. Приведем пример для случая, когда v распределена по равномерному закону в диапазоне от −1,0 до 1,0. На рисунке для $\theta = 3$, $n = 100$ и при $\sigma_\varepsilon = 0$, $\sigma_\eta = 1$ приведены типичные решения при использовании МНК, МНР и предложенного способа. Там же приведены истинная прямая и точки выборки.

3. Приведем пример для случая, когда v является СВ с несимметричным распределением.

Таблица 2

Расчет параметра связи при случайной входной величине, распределенной по логнормальному закону

Параметры ошибок	Параметры оценки	Использование симметрии ПВ		МНР	Вычисление по формуле (23)	Вычисление по формуле (20)	Вычисление по формуле (24)
		С использованием формулы (19)	Совместный выбор углов				
Объем выборки $n = 500$							
$\sigma_\varepsilon = 1{,}2$, $\sigma_\eta = 0{,}3$	$\bar{\theta}$	1,0451	1,0484	1,3239	1,0390	1,0381	1,0497
	σ_θ	0,0721	0,0489	0,0565	0,1016	0,1069	0,0596
	m_{e2}	0,0052	0,0024	0,0782	0,0104	0,0116	0,0036
$\sigma_\varepsilon = 0{,}4$, $\sigma_\eta = 1{,}3$	$\bar{\theta}$	1,0423	1,0470	0,8001	0,9648	1,0678	1,0501
	σ_θ	0,0663	0,0533	0,0347	0,1171	0,1344	0,0688
	m_{e2}	0,0045	0,0028	0,0638	0,0210	0,0184	0,0047
Объем выборки $n = 50$							
$\sigma_\varepsilon = 1{,}2$, $\sigma_\eta = 0{,}3$	$\bar{\theta}$	1,1481	1,1261	1,3266	1,1735	1,1421	1,0076
	σ_θ	0,2494	0,1298	0,1413	0,7650	0,4424	0,3451
	m_{e2}	0,0723	0,0227	0,0963	0,6010	0,2041	0,9210
$\sigma_\varepsilon = 0{,}4$, $\sigma_\eta = 1{,}3$	$\bar{\theta}$	0,9715	1,0030	0,8307	1,2061	1,0911	1,0122
	σ_θ	0,1871	0,1449	0,1328	1,8382	0,4910	1,7837
	m_{e2}	0,0358	0,0232	0,0657	3,3990	0,2435	3,1828

В работе [4] в модельном примере использовался логнормальный закон. Для сравнительной оценки методов будем также использовать его. Был проведен расчет для нормальных СВ ε и η с СКО, соответственно, 1,2 и 0,3; 0,4 и 1,3 и СВ v, распределенной по логнормальному закону LN(1; 0,5). При этом брались значения объема выборки $n = 500$ и $n = 50$, коэффициент $\theta = 1{,}05$. Логнормальный закон относится к распределениям с тяжелыми хвостами, поэтому число экспериментов для $n = 50$ брали равным 500 для всех способов вычислений, кроме использования симметрии плотностей с независимым выбором углов, в этом случае число повторений было равно 100. В отличие от предыдущего примера здесь должны сравниваться характеристики сглаженной ядерным способом оценки $f_{exy}(x_0, y_0)$ лишь относительно оси Ox_0. Значения дисперсии по обеим координатам плотности нормального распределения, используемой в качестве ядра, принимались равными 2,25. Результаты расчетов сведены в табл. 2.

Из табл. 2 видно, что в предлагаемом способе с использованием симметрии плотности значение второго момента ошибки в основном определяется дисперсией оценки, а в МНР — смещением оценки.

Заключение

Метод с использованием симметричности плотностей гауссовых помех имеет превосходство над известными методами по точности. Возможно распространение предложенной методики на многомерный случай.

Список литературы

1. **Жилинская Е. И., Товмаченко Н. Н., Федоров В. В.** Методы регрессионного анализа при наличии ошибок в предикторных переменных. М.: АН СССР, Научный совет по комплексной программе "Кибернетика", 1979. С. 16—25.

2. **Айвазян С. А., Енюков И. С., Мешалкин Л. Д.** Прикладная статистика: Исследование зависимостей. М.: Финансы и статистика, 1985. С. 235—244.

3. **Зайцев В. Г.** К использованию ядерных оценок при сглаживании данных // Заводская лаборатория. Диагностика материалов. 2017. Т. 83, № 5. С. 66—71.

4. **Тимашев С. А., Тырсин А. Н.** Оценивание линейных структурных соотношений между случайными величинами // Заводская лаборатория. Диагностика материалов. 2010. Т. 76, № 3. С. 68—71.

5. **Gillard J. W.** Method of moments estimation in linear regression with errors in both variables // Communications in Statistics: Theory and Methods. 2014. Vol.43, N. 15. P. 3208—3222. URL: http://orca.cf.ac.uk/71432/.

V. G. Zaycev, Lead Engineer, e-mail: zaycev@looch.ru,
Scientific Production Enterprise of Geophysical Equipment "Looch"
Novosibirsk, 630051, Russian Federation

To Evaluation of Parameters of Models at Presence of Errors in Entrance and Output Variables. Linear Case

This paper presents the ways to identify systems described by models with errors in recording of input and output variables. It is considered that in the absence of information on the error parameters in the case of their large values, it is impossible to obtain acceptable estimates of all the required parameters of the model without additional, in comparison with the regression analysis, assumptions. The determination of the regression parameters in the presence of errors in the input and output variables is important for solving many tasks related to data processing. In all existing approaches, either the assumptions regarding the noise in the form of e.g. given relations between the noises dispersions are made, or the method of successive approximations is applied. To determine the parameters of the linear function, it is proposed to use the condition of the symmetry of the joint probability density of the observed input and output variables in the oblique coordinates. For the case of normality of the noise, the article gives a formula for determining the parameter of the relationship via the estimates of the fourth-order semi-invariants.

Keywords*: structural analysis, symmetric distribution of hindrances, oblique-angled system of coordinates, solvency of estimation*

DOI: 10.17587/it.26.555-563

References

1. **Zhilinskaya E. I., Tovmachenko I. S., Feodorov V. V.** Methods for regression analysis with errors in the predictor variables, Moscow, Nauka Publishers, 1979, 34 p. (in Russian).

2. **Aivazyan S. A., Yenukov I. S., Mechalkin L. D.** Applied statistics: Research of dependences, Moscow, Finance And Statistics Publishers, 1985, 487 p. (in Russian).

3. **Zaycev V. G.** On the Use of Kernel Estimators in Data-Smoothing, *Zavodskaya laboratoriya. Diagnostica materialov*, 2017, vol. 83, no. 5, pp. 66—71 (in Russian).

4. **Timachov S. A., Tyrsin A. N.** Estimation of linear structural relationships between the random variables, *Zavodskaya laboratoriya. Diagnostika materialov*, 2010, vol. 76, no. 3, pp. 68—71 (in Russian).

5. **Gillard J. W.** Method of moments estimation in linear regression with errors in both variables. *Communications in Statistics: Theory and Method,* 2014, vol. 43, no. 15, pp. 3208—3222, available at: http://orca.cf.ac.uk/71432/

УДК 621.391; 519.688 DOI: 10.17587/it.26.563-569

А. Ю. Спасёнов, аспирант, ассистент кафедры, e-mail: a.spasenov@mail.ru,
К. В. Кучеров, аспирант, ассистент, e-mail: cvkucherov@yandex.ru,
Т. М. Волосатова, канд. техн. наук, доц., e-mail: tamaravol@gmail.com,
Д. М. Жук, канд. техн. наук, доц., e-mail: zhuk_d@mail.ru,
Московский государственный технический университет имени Н. Э. Баумана
(национальный исследовательский университет)

Оценка состояния сложных технических объектов с использованием структурно-модального анализа квазипериодических временных рядов

Представлен способ описания и оценки технических состояний сложного динамического объекта с использованием метода структурно-модального анализа многомерных временных рядов. Показана возможность автоматической оценки динамики изменения состояния объекта и получения диагностической информации на основе портрета технического состояния объекта. Предлагаемый подход может быть использован для создания специального математического обеспечения, направленного на автоматический анализ состояния сложных технических систем.

Ключевые слова: *анализ временных рядов, техническое состояние объекта, модальный анализ, тематическое моделирование*

Введение

В различных областях науки и техники встречаются системы, моделирование которых представляет собой сложную задачу в связи с наличием глубоких зависимостей между их составными частями. Такие системы, как правило, нелинейны, гетерогенны и могут иметь

обратные связи [1—3]. Для описания систем подобного рода в настоящее время принято выделять класс сложных технических систем (СТС).

СТС, как правило, эксплуатируются совместно с системами мониторинга, что позволяет снизить возможные риски от возникновения разладки в системе. При этом задача создания систем мониторинга усложняется в той же мере, в которой конструктивно усложняются СТС. Наиболее важной задачей системы мониторинга является работа в режиме реального времени (жесткость режима зависит от конкретной решаемой задачи) или, по крайней мере, в близком к нему по латентности. При обеспечении такого режима работы качество системы мониторинга может быть значительно улучшено за счет использования совместно с информацией о текущем состоянии объекта исторических наблюдений, хранящихся в виде системы знаний.

Обычно СТС имеют конечное число физически интерпретируемых состояний и в один момент времени могут находиться в каком-либо одном из них. Большая часть возможных состояний выделяется на основе априорной информации о СТС, и такие состояния имеют наибольшее значение для эксплуатации и обслуживания СТС. В связи с этим актуальной является задача отнесения текущего состояния СТС к одному из известных. Существуют работы [4], в рамках которых при использовании диагностических процедур и априорной информации о структуре исследуемой системы эта задача может быть решена с определенной степенью точности.

Исходные данные для анализа СТС часто представлены в виде квазипериодического временного ряда (ВР), каждый элемент которого представляет собой набор характерных признаков (ХП), описывающих состояние СТС в определенный момент времени [5]. Повышение точности классификации состояния СТС связано в первую очередь с исследованием применимости методов поиска характерных участков, периодических участков и трендов во ВР. Особый интерес представляет последующая обработка найденных участков, состоящая в извлечении ХП с последующим описанием принадлежности системы к конкретному состоянию с использованием методов тематического моделирования. Такой подход позволит значительно сократить число обрабатываемых сегментов ВР и даст возможность оценивать динамику изменения ХП только специфических фрагментов записи. Целью данной работы является демонстрация возможностей метода структурно-модального анализа квазипериодических многомерных ВР для оценки состояния СТС.

Постановка задачи структурно-модального анализа квазипериодических временных рядов

Одним из примеров сигнала реальной СТС является сигнал вибрации, регистрируемый на станке при обработке изделий (рис. 1). Обработка происходит в течение нескольких циклов, один из которых представлен в увеличенном виде на рис. 1.

Анализ вибраций станка выполняется в целях ранней диагностики поломок и предотвращения порчи заготовок деталей. При этом необходимо, чтобы система управления станком была способна анализировать изменения состояния СТС и в случае перехода в состояние (состояния) "поломки" могла оперативно реагировать на это.

Использование для решения этой задачи методов структурного анализа квазипериодических нестационарных сигналов осложняется большой размерностью исходных данных — получение близких к требуемым характеристик латентности при таком подходе невозможно. Однако следует отметить, что информативными являются не все участки анализируемого сигнала, а также тот факт, что на основе априорных знаний информативные участки могут быть отнесены к заранее известному конечному множеству интерпретируемых состояний.

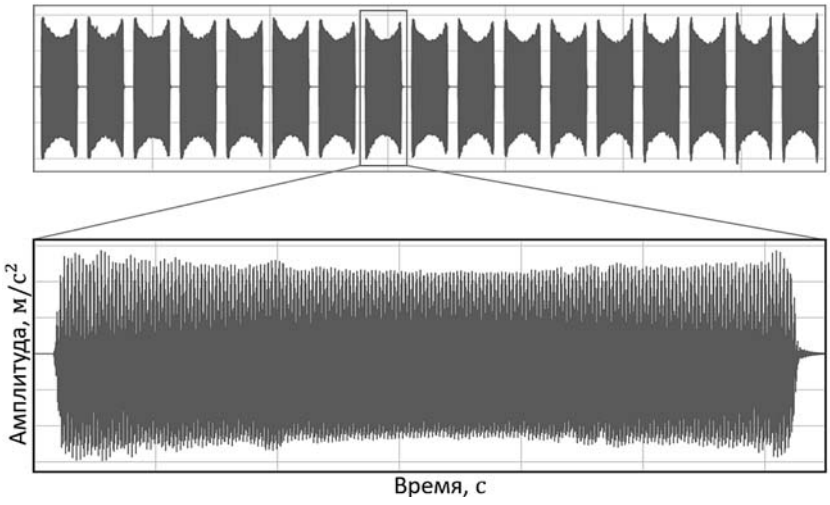


Рис. 1. Сегментация сигнала вибрации

Исходя из этих соображений в настоящей работе предлагается оригинальный подход к структурно-модальному анализу квазипериодических ВР. На первом этапе предлагаемого подхода происходит выделение паттернов в сигнале с использованием методов сегментации ВР. Далее выполняется параметрическое описание процессов, протекающих в выявленных сегментах сигнала, и создание вероятностного портрета состояния системы.

Сегментация квазипериодических временных рядов

Динамика СТС, находящейся в определенном состоянии, характеризуется участком (сегментом) исходного ВР. В рамках такого описания задача сегментации представляет собой задачу определения факта перехода между состояниями. Она может быть решена лишь в том случае, если система обладает фундаментальным свойством наблюдаемости, означающим, что по информации на выходе можно полностью восстановить информацию о состояниях системы.

Благодаря наличию априорной информации о состояниях, представленных в реальном мире СТС, решение задачи сегментации состоит в нахождении преобразования, пригодного для сопоставления ВР с одним или несколькими характерными для СТС в определенном состоянии участками (паттернами). Задача сегментации может быть представлена в виде композиции двух подзадач. Первой из них является поиск подобия участков сигнала, при этом длина участков может быть постоянной или переменной. Второй задачей является локализация во времени внезапных изменений динамики системы, формально задаваемой в виде зависимости элементов ВР от индекса.

При анализе сигналов реальных систем важно выбрать адекватную модель, учитывающую особенности анализируемых сигналов и позволяющую интерпретировать получаемые результаты с физической точки зрения. Нестационарность анализируемых ВР не позволяет применять к ним классические методы частотного анализа, так как спектральные характеристики таких ВР меняются со временем.

Общепринятым подходом к анализу нестационарных ВР является использование оконных преобразований с временным окном фиксированной длины. При этом предполагается, что ширина окна может быть подобрана таким образом, что участок ряда внутри него будет квазистационарным. Однако получаемые с реальных систем сигналы, как правило, не обладают свойством локальной стационарности. Это обусловлено нелинейным характером динамики системы, что выражается присутствием в анализируемом ВР быстро затухающих мод, быстрых локальных изменений значений модальных параметров или резких изменений структуры системы. Таким образом, применение методов сегментации с фиксированным размером окна не всегда возможно из-за свойств анализируемого ВР. Более представительной [6] является модель локально-переходных ВР или модель локально-переходных изменений [7, 8]. Она характеризует временные интервалы, в которых нарушается локальная стационарность ВР.

Можно выделить ряд направлений, в рамках которых на данный момент времени были получены наилучшие результаты в решении задачи сегментации нестационарных ВР: генетические алгоритмы [9, 10], многомасштабный корреляционный анализ [11—13], методы модальной декомпозиции [14—16], вейвлет-анализ [17], применение шейплетов [18, 19].

Для решения задач первого этапа структурно-модального анализа квазипериодических ВР может быть выбран любой из перечисленных методов с учетом ограничений и особенностей применения в рамках конкретной предметной области.

Оценка состояния сложных технических объектов

Основным фактором, связывающим сигналы определенного вида со специфическими для них процедурами обработки, является выбор формального математического описания реальных данных и измерений. Усложнение природы исследуемых сигналов требует совершенствования моделей и методов их обработки. Происходит разделение на параметрические, непараметрические и полупараметрические методы обработки в зависимости от сложности представления исследуемого процесса [5].

Задача контроля и мониторинга технического состояния объектов относится к задачам общей теории распознавания образов текущего состояния объектов [5]. Решение этой задачи в технической диагностике основано на диагностических моделях, устанавливающих связь между состоянием технической системы

и его отображением в пространстве диагностических признаков.

Состояние многих динамических объектов с распределенными параметрами может быть представлено рядами следующего вида [20]:

$$Q(x,t) = \sum_{k=1}^{\infty} a_k \gamma_k(t) \varphi_k(x,y,z),$$

где $\varphi_k(x, y, z)$ и $\gamma_k(t)$ — пространственная и временная моды соответственно. Числа a_k зависят от внешних входных воздействий (внешних возмущений, начальных условий и т.д.). Функции $\varphi_k(x, y, z)$ и $\gamma_k(t)$ являются внутренней характеристикой распределенной системы.

Модальные параметры, описывающие протекающие процессы в интересующих сегментах, будут являться их ХП. Хорошо зарекомендовавшим себя методом параметрического модального анализа сигналов является метод Прони [21—23]. Данный метод имеет высокую разрешающую способность и позволяет определить особенности временной эволюции динамических процессов.

Для решения задачи анализа технического состояния объектов на основе полученных диагностических признаков воспользуемся моделью [4], которая представляет собой два упорядоченных множества. Первое множество является моделью объекта анализа, второе — моделью процесса определения технического состояния объекта, т. е. процесса анализа:

$$\begin{aligned} M_{\text{о}} &= \langle S, \Pi, \Sigma, P, \Phi \rangle, \\ M_{\text{п}} &= \left\langle S, \Omega, P, \widehat{\Pi} \right\rangle, \end{aligned} \qquad (1)$$

где $M_{\text{о}}$ — модель объекта анализа; $M_{\text{п}}$ — модель процесса определения состояния объекта, т. е. процесса анализа; $S = \{S_i | i = \overline{1,m}\}$ — множество технических состояний, в одном из которых может находиться проверяемый объект; $\widehat{\Pi} = \{\hat{\pi}_j | j = \overline{1,n}\}$ — множество проверок, взаимно однозначно соответствующее множеству $\Pi = \{\pi_j | j = \overline{1,n}\}$ диагностических признаков, на котором все технические состояния $S_i \in S$ попарно различимы; $\Sigma = \{\sigma_{ij} | i = \overline{1,m}, j = \overline{1,n}\}$ — множество модельных значений признаков, каждый из которых означает наиболее вероятностный исход проверки $\hat{\pi}_j \in \widehat{\Pi}$ в ТС $S_i \in S$; $P = \left\{ P(S_i) | \sum_{i=1}^{m} P(S_i) = 1 \right\}$ — множество вероятностей ТС $S_i \in S$; $\Omega = \{R | R \subseteq S\}$ — алгебра подмножеств множества S, в которой элементы R имеют смысл информационных состояний моделируемого процесса; $\Phi: S \times \Pi \to \Sigma$ — отображение, устанавливающее связь между элементами множеств Σ, S и Π, согласно которому $\sigma_{ij} = \Phi(S_i, \hat{\pi}_j)$, $\hat{\pi}_j \in \widehat{\Pi}$, $S_i \in S$.

В результате вероятностно-динамической модели реализуется последовательная процедура оценки технического состояния. Для уменьшения вычислительной сложности алгоритмов параметрического модального анализа предлагается использовать оконное преобразование. На рис. 2 показано разбиение одного сегмента записи на отдельные фрагменты с использованием оконной функции. Варьируемыми параметрами данного шага являются ширина окна и сдвиг.

Каждый полученной фрагмент подается на вход методу параметрического модального анализа Прони. Полагая, что наблюдаемые данные $x[n]$ имеют N комплексных отсчетов $x[1]$, $x[2]$, ..., $x[N]$, метод Прони будет сопоставлять эти данные с суммой комплексных функций:

$$y[n] = \sum_{k=1}^{M} A_k \mathbf{e}^{(n-1)(\alpha_k + j2\pi f_k)\Delta + j\theta_k}$$

для $n = 1, 2, ..., N$, где $j^2 = -1$, а Δ — интервал дискретизации. Объектами оценивания являются амплитуда комплексных экспонент A_k, па-

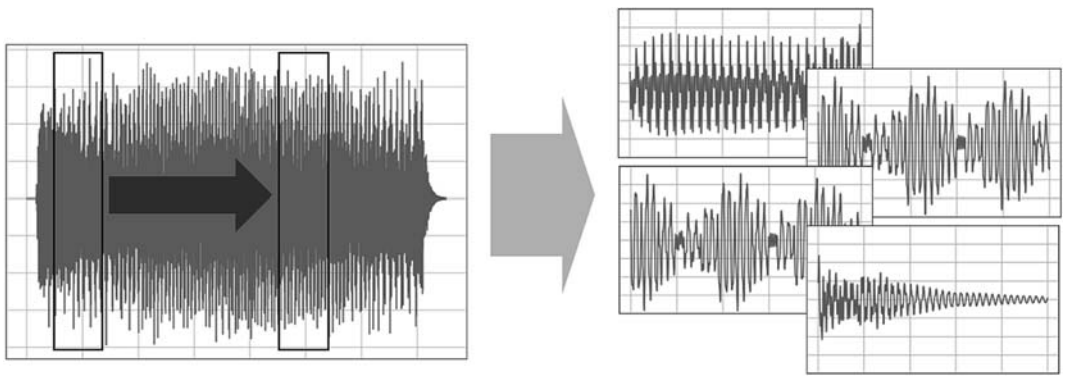


Рис. 2. Оконный анализ

раметр затухания α_k, гармоническая частота f_k и фаза θ_k. Значения этих параметров являются ХП для конкретного уровня разложения сигнала, представляющими собой диагностические признаки в модели (1). Если эти параметры определены корректно, степень приближения исходного сигнала будет высокой [21]. На рис. 3 (см. вторую сторону обложки) представлено параметрическое модальное разложение сигнала в разных диапазонах частот (F_1, F_2, F_3, ..., F_N) и соответствующий им набор ХП.

Фрагменту записи при этом ставится в соответствие совокупность декомпозиций, причем каждый найденный набор модальных параметров может рассматриваться как терм, а их комплекс — как текстовый документ [24]. Таким образом, в соответствие анализируемому ВР ставится корпус документов. В случае наличия априорной информации о частотных составляющих сигнала можно жестко установить все возможные диапазоны значений модальных параметров и выполнить процедуру категоризации значений ХП. Если свойства исследуемого процесса заранее неизвестны, то можно воспользоваться методами кластеризации данных. Число кластеров и параметры работы алгоритмов будут являться настраиваемыми параметрами предлагаемого метода. В результате будет получено множество W всех возможных термов. Формирование словаря можно понимать как выделение набора макрособытий в развитии исследуемого процесса [25]. На рис. 4 (см. вторую сторону обложки) представлен результат получения портрета состояния системы в рамках характерного фрагмента исходного сигнала. Под портретом понимается распределение частот встречаемости наборов модальных параметров в рамках одного фрагмента ВР.

При анализе СТС сложно заранее определить все возможные состояния исследуемого объекта. Для поиска возможных состояний системы можно воспользоваться алгоритмами тематического моделирования на основе полученного корпуса документов, осуществляющими мягкую кластеризацию путем разделения документов между несколькими кластерами. Тематическое моделирование обладает существенным запасом гибкости, позволяющим обрабатывать сложно структурированные данные и применять тематический анализ совместно с другими методами анализа текстов [26]. Согласно формуле полной вероятности и гипотезе условной независимости, распределение термов в документе описывается вероятностной смесью распределений термов в темах:

$$p(w|d) = \sum_{t \in T} p(w|t,d)\, p(t|d) = \\ = \sum_{t \in T} p(w|t)\, p(t|d) \sum_{t \in T} \varphi_{wt} \theta_{td},$$

где w — терм; d — документ, связанный с темой t, $\varphi_{wt} = (p|t)$ — вероятностная смесь распределений термов в темах с весами $\theta_{td} = p(t|d)$. Каждую полученную тему можно интерпретировать как возможное состояние, в котором может находиться исследуемый объект.

На рис. 5 и 6 (см. вторую сторону обложки) представлены результаты сжатия представления исследуемого процесса в три характерных процесса с использованием модели латентного размещения Дирихле [27, 28], являющейся одной из наиболее популярных моделей тематического моделирования. Радиус окружностей на изображениях определяется долей наборов модальных параметров в сегменте анализируемого ВР, которые содержат соответствующие значения амплитуды и частоты. Каждый полученный фрагмент определяется свойствами процессов, протекающих во время обработки изделия.

Заключение

В работе описан оригинальный метод структурно-модального анализа квазипериодических ВР. Предлагаемый метод может быть использован для решения задачи диагностики и мониторинга состояния СТС в различных прикладных областях. Описанный метод позволяет выявлять динамику изменения состояния систем в автоматическом или полуавтоматическом режимах в зависимости от объема априорной информации об анализируемой СТС. Комбинирование методов интеллектуальной обработки данных, применяемых в различных областях, позволяет получить интерпретируемые результаты в целях улучшения качества решения поставленных задач оценки СТС.

Дальнейшее развитие данного подхода возможно за счет формирования словаря состояний с термами, отвечающими за характер изменения параметров ВР, и использования портрета состояний системы для решения различных задач интеллектуального анализа данных.

Список литературы

1. **Bar-Yam Y.** General Features of Complex Systems // Knowledge management, organizational intelligence and learning, and complexity. 2002. Vol. 1, N. 1. P. 3—13.

2. **Цветков В. Я.** Сложные технические системы // Образовательные ресурсы и технологии. 2017. № 3. С. 86—91.

3. **Кудж С. А.** Многоаспектность рассмотрения сложных систем // Перспективы науки и образования. 2014. № 1. С. 38—43.

4. **Копкин Е. В., Кобзарев И. М., Зверева Е. Е.** Квазиоптимальный алгоритм построения гибкой программы анализа технического состояния объекта // Наукоемкие технологии в космических исследованиях Земли. 2017. Т. 9, № 3. С. 4—12.

5. **Лоскутов А. И., Козырев Г. И., Клыков В. А., Шестопалова О. Л.** Синтез адаптивных математических моделей бортовых радиоэлектронных систем космических аппаратов на основе применения гомологичных математических структур // Труды СПИИРАН. 2018. № 1. С. 169—194.

6. **Кухаренко Б. Г.** Исследование по методу Прони динамики систем на основе временных рядов // Труды Московского физико-технического института. 2009. Т. 1, № 2. С. 176—192.

7. **Porat B., Friedlander B.** Performance analysis of a class of transient detection algorithms: a unified framework // IEEE Transactions on Signal Processing. 1992. Vol. 40, N. 10. P. 2536—2545.

8. **Thornburg H., Gouyon F**. A flexible analysis-synthesis method for transients // Proceedings of International Computer Music Conference. Berlin, 2000. P. 7—11.

9. **Azami H., Mohammadi K., Hassanpour H.** An improved signal segmentation method using genetic algorithm // International Journal of Computer Applications. 2011. Vol. 29, N. 8. P. 5—9.

10. **Krajca V., Petranek S., Patakova I., Varri A.** Automatic Identification of Significant Graphoelements in Multichannel EEG Recordings by Adaptive Segmentation and Fuzzy Clustering // International Journal of Biomedical Engineering. 1991. Vol. 28, N. 1. P. 71—89.

11. **Анциперов В. Е.** Обнаружение ритмов головного мозга человека на основе корреляции аналитических спектров ЭЭГ в основных диапазонах частот // Журнал радиоэлектроники. 2014. № 5. С. 13—24.

12. **Анциперов В. Е.** Оценивание характера последействия случайных точечных процессов методами многомасштабного корреляционного анализа // Журнал радиоэлектроники. 2015. № 6. С. 12—32.

13. **Анциперов В. Е.** Многомасштабный корреляционный анализ нестационарных, содержащих квазипериодические участки сигналов // Радиотехника и электроника. 2008. № 53. С. 73—85.

14. **Kumaresan R., Tufts D. W.** Estimating the parameters of exponentially damped sinusoids and pole-zero modeling in noise // IEEE Transactions on Acoustics, Speech, and Signal Processing. 1982. Vol. 30, N. 4. P. 833—840.

15. **Tufts D. W., Kumaresan R.** Singular value decomposition and improved frequency estimation using linear prediction // IEEE Transactions on Acoustics, Speech, and Signal Processing. 1982. Vol. 30, N. 4. P. 671—675.

16. **Hua Y., Sarkar T. K.** On SVD for estimating generalized eigenvalues of singular matrix pencil in noise // IEEE Transactions on Acoustics, Speech, and Signal Processing. 1991. Vol. 39, N. 4. P. 892—900.

17. **Назимов А. И.** и др. Распознавание осцилляторных паттернов на электроэнцефалограмме на основе адаптивного вейвлет-анализа // Вестник Тамбовского университета. Серия: Естественные и технические науки. 2013. Т. 18, № 1. С. 10—14.

18. **Карпенко А. П., Кострубин М. С., Чернышев А. С**. Эффективность классификации многомерных временных рядов с помощью шейплетов // Машиностроение и компьютерные технологии. 2015. № 11. С. 382—405.

19. **Карпенко А. П., Сотников П. И.** Модифицированный метод классификации многомерных временных рядов с использованием шейплетов // Вестник Московского государственного технического университета им. НЭ Баумана. Серия "Приборостроение". 2017. № 2. С. 46—65.

20. **Бутковский А. Г.** Структурная теория распределенных систем. М.: Наука, 1977. 320 с.

21. **Marple S. L.** Digital Spectral Analysis with Applications. NY.: Prentice Hall, 1987. 492 p.

22. **Priyanka S. Pariyal, Dhara M. Koyani, Daizy M. Gandhi, Sunil F. Yadav, Dharam J. Shah, Ankit Adesara.** Comparison based Analysis of Different FFT Architectures // International Journal of Image, Graphics and Signal Processing(IJIGSP). 2016. Vol. 8, N. 6. P. 41—47.

23. **Досько С. И., Волосатова Т. М., Спасенов А. Ю., Кучеров К. В.** Частотно-временной анализ биомедицинских сигналов на основе методов Хуанга, Прони, Фурье // Динамика сложных систем. 2020. Т. 14, № 1. С. 32—38.

24. **Жук Д. М., Волосатова Т. М., Спасенов А. Ю., Кучеров К. В.** Оценка динамических систем с использованием модального-лингвистического анализа многомерных временных рядов // Динамика сложных систем. 2020. Т. 14, № 1. С. 38—45.

28. **Браверман Э. М., Мучник И. Б.** Структурные методы обработки эмпирических данных. М.: Наука, 1983. 464 с.

26. **Воронцов К. В., Потапенко А. А.** Регуляризация, робастность и разреженность вероятностных тематических моделей // Компьютерные исследования и моделирование. 2012. Т. 4, № 4. С. 693—706.

27. **Blei D. M., Ng A. Y., Jordan M. I.** Latent Dirichlet allocation // Journal of Machine Learning Research. 2003. Vol. 3, N. 1. P. 993—1022.

28. **Potapenko A. A., Vorontsov K. V.** Robust PLSA Performs Better Than LDA // European Conference on Information Retrieval ECIR-2013. Moscow, 2013. P. 784—787.

A. Yu. Spasenov, Assistant, Postgraduate Student,
Department "Systems of the Automated Designing", e-mail: a.spasenov@mail.ru,
K. V. Kucherov, Assistant, Postgraduate Student,
Department "Computer Systems, Complexes and Networks", e-mail: Cvkucherov@yandex.ru,
T. M. Volosatova, Ph.D., Associate Professor,
Department "Systems of the Automated Designing", e-mail: tamaravol@gmail.com,
D. M. Zhuk, Ph.D., Associate Professor,
Department "Systems of the Automated Designing", e-mail: zhuk_d@mail.ru,
Bauman Moscow State Technical University, Moscow, Russian Federation

Analysis of Quasi-Periodic Time Series by a Structural-Modal Method for Monitoring and Diagnostics of Complex Technical Systems

The possibilities of using the structural-modal method for monitoring and diagnostics the technical states of a complex technical systems are presented. The main idea of the proposed method is to combine the methods of parametric modal decomposition of signals and thematic modeling methods used for soft clustering of attributes of segments of time series. The initial data for analysis are often presented as a quasiperiodic time series. Each period of the time series is a set of modal parameters. The first stage of the proposed approach is time series segmentation. Second stage is a parametric modal decomposition of the processes occurring in the identified signal segments. The final stage is creation of a probabilistic portrait of each state of the system. The structural-modal method allows to identify the dynamics of state of systems changes in automatic or semi-automatic modes, depending on the amount of a priori information about the analyzed complex technical system. The proposed method can be used to solve the problem of diagnosing and monitoring the condition of complex technical system in various application areas. The proposed approach can be used to create special mathematical software aimed at the automatic analysis of the state of complex technical systems.

Keywords: *time series analysis, technical state, modal analysis, thematic modeling*

DOI: 10.17587/it.26.563-569

References

1. **Bar-Yam Y.** General Features of Complex Systems, *Knowledge Management, Organizational Intelligence and Learning, and Complexity*, 2002, vol. 1, no. 1, pp. 3—13.

2. **Cvetkov V. Ja.** Complex technical systems, *Obrazovatel'nye Resursy i Tehnologii*, 2017, no. 3, pp. 86—91 (in Russian).

3. **Kudzh S. A.** The multidimensional nature of the consideration of complex systems, *Perspektivy Nauki I Obrazovanija*, 2014, no. 1, pp. 38—43 (in Russian).

4. **Kopkin E. V., Kobzarev I. M., Zvereva E. E.** Quasi-optimal algorithm for constructing a flexible program for analyzing the technical condition of an object, *Naukoemkie Tehnologii v Kosmicheskih Issledovanijah Zemli*, 2017, vol. 9, no.3, pp. 4—12 (in Russian).

5. **Loskutov A. I., Kozyrev G. I., Klykov V. A., Shestopalova O.L.** Synthesis of adaptive mathematical models of on-board electronic systems of spacecraft based on the use of homologous mathematical structures, *Trudy SPIIRAN*, 2018, no. 1, pp. 169—194 (in Russian).

6. **Kuharenko B. G.** Research on the Proni method of system dynamics based on time series, *Trudy Moskovskogo fiziko-tehnicheskogo instituta,* 2009, vol. 1, no. 2, pp. 176—192 (in Russian).

7. **Porat B., Friedlander B.** Performance analysis of a class of transient detection algorithms: a unified framework, *IEEE Transactions on Signal Processing*, 1992, vol. 40, no. 10, pp. 2536—2545.

8. **Thornburg H., Gouyon F.** A flexible analysis-synthesis method for transients, *Proceedings of International Computer Music Conference*, Berlin, 2000, pp. 7—11.

9. **Azami H., Mohammadi K., Hassanpour H.** An improved signal segmentation method using genetic algorithm, *International Journal of Computer Applications,* 2011, vol. 29, no. 8, pp. 5—9.

10. **Krajca V., Petranek S., Patakova I., Varri A.** Automatic Identification of Significant Graphoelements in Multichannel EEG Recordings by Adaptive Segmentation and Fuzzy Clustering, *International Journal of Biomedical Engineering*, 1991, vol. 28, no. 1, pp. 71—89.

11. **Anciperov V. E.** Detection of human brain rhythms based on the correlation of analytical EEG spectra in the main frequency ranges, *Zhurnal Radiojelektroniki*, 2014, no. 5, pp. 13—24 (in Russian).

12. **Anciperov V. E.** Estimation of the nature of the aftereffect of random point processes using multiscale correlation analysis, *Zhurnal Radiojelektroniki*, 2015, no. 6, pp. 12—32 (in Russian).

13. **Anciperov V. E.** Multiscale correlation analysis of non-stationary, containing quasiperiodic sections of signals, *Radiotehnika i jelektronika*, 2008, no. 53, pp. 73—85 (in Russian).

14. **Kumaresan R., Tufts D. W.** Estimating the parameters of exponentially damped sinusoids and pole-zero modeling in noise, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 1982, vol. 30, no.. 4, pp. 833—840.

15. **Tufts D. W., Kumaresan R.** Singular value decomposition and improved frequency estimation using linear prediction, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 1982, vol. 30, no. 4, pp. 671—675.

16. **Hua Y., Sarkar T. K.** On SVD for estimating generalized eigenvalues of singular matrix pencil in noise, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 1991, vol. 39, no. 4, pp. 892—900.

17. **Nazimov A. I.** Recognition of oscillatory patterns on an electroencephalogram based on adaptive wavelet analysis, *Vestnik Tambovskogo universiteta. Serija: Estestvennye i tehnicheskie nauki*, 2013, vol. 18, no. 1, pp. 10—14 (in Russian).

18. **Karpenko A. P., Kostrubin M. S., Chernyshev A. S.** The effectiveness of the classification of multidimensional time series using shapeplates, *Mashinostroenie i Komp'juternye Tehnologii*, 2015, no. 11, pp. 382—405 (in Russian).

19. **Karpenko A. P., Sotnikov P. I.** A modified method for the classification of multidimensional time series using shapeplates, *Vestnik Moskovskogo gosudarstvennogo tehnicheskogo universiteta im. N. Je. Baumana. Serija "Priborostroenie",* 2017, no. 2, pp. 46—65 (in Russian).

20. **Butkovskij A. G.** Structural theory of distributed systems, Moscow, Nauka, 1977, 320 p. (in Russian).

21. **Marple S. L.** Digital Spectral Analysis with Applications, NY., Prentice Hall, 1987, 492 p.

22. **Priyanka S. Pariyal, Dhara M. Koyani, Daizy M. Gandhi, Sunil F. Yadav, Dharam J. Shah, Ankit Adesara.** Comparison based Analysis of Different FFT Architectures, *International Journal of Image, Graphics and Signal Processing(IJIGSP)*, 2016, vol. 8, no. 6, pp. 41—47.

23. **Dos'ko S. I., Volosatova T. M., Spasjonov A. Ju., Kucherov K. V.** Frequency-time analysis of biomedical signals based on the methods of Huang, Proni, Fourier, *Dinamika Slozhnyh Sistem*, 2020, vol. 14, no. 1, pp. 32—38 (in Russian).

24. **Zhuk D. M., Volosatova T. M., Spasjonov A. Ju., Kucherov K. V.** Evaluation of dynamic systems using modal linguistic analysis of multidimensional time series. The dynamics of complex systems, *Dinamika Slozhnyh Sistem,* 2020, vol. 14, no. 1, pp. 38—45 (in Russian).

23. **Braverman Je. M., Muchnik I. B.** Structural methods for processing empirical data, Moscow, Nauka, 1983, 464 p. (in Russian).

26. **Voroncov K. V., Potapenko A. A.** Regularization, robustness and sparseness of probabilistic thematic models, *Komp'juternye Issledovanija I Modelirovanie,* 2012, vol. 4, no. 4, pp. 693—706 (in Russian).

27. **Blei D. M., Ng A. Y., Jordan M. I.** Latent Dirichlet allocation, *Journal of Machine Learning Research*, 2003, vol. 3, no. 1, pp. 993—1022.

28. **Potapenko A. A., Vorontsov K. V.** Robust PLSA Performs Better Than LDA, *European Conference on Information Retrieval ECIR-2013*, Moscow, 2013, pp. 784—787.

УДК 004.36 DOI: 10.17587/it.26.570-576

С. А. Инютин, д-р техн. наук, профессор,e-mail: inyutin_int@mail.ru,
Московский авиационный институт (национальный исследовательский университет) (МАИ)

Метрики в модулярном векторном пространстве

Модулярные представления числовых величин в форме кортежей вычетов по взаимно простым модулям из некоторой конечной совокупности можно рассматривать как множество модулярных векторов. Множество модулярных векторов рассматривается как линейное подпространство в векторном пространстве, содержащем векторы с компонентами, имеющими ограниченное значение, его кольцевая структура не учитывается. Анализируются свойства линейного модулярного подпространства, способы определения новых скалярных произведений, что позволяет ввести новые метрики. Введенные алгебраические конструкции предназначены для анализа сходимости многоразрядных параллельных вычислительных процессов в больших компьютерных диапазонах, оперирующих с числовыми величинами в модулярном представлении. Введенные числовые векторные представления ориентированы на применение в модулярных реконфигурируемых вычислительных системах SIMD-архитектуры.

Ключевые слова: *модулярное векторное подпространство, модулярные скалярные произведения, модулярные метрики, параллельный модулярный вычислительный процесс, многопроцессорная реконфигурируемая система*

Введение

В модулярной компьютерной системе счисления числовые величины $A(\mathrm{mod}\,P) \in \{0,1,2,\ldots P-1\}$, принадлежащие полной системе вычетов по модулю P, представлены векторами с компонентами — наименьшими неотрицательными вычетами (абсолютно наименьшими) по простым (взаимно простым) модулям:

$$A(\mathrm{mod}\,P) \leftrightarrow \\ \leftrightarrow (a_1(\mathrm{mod}\,p_1),\ldots, a_i(\mathrm{mod}\,p_i),\ldots a_n(\mathrm{mod}\,p_n)), \quad (1)$$

где $a_i \equiv A(\mathrm{mod}\,p_i), a_i \in \{0,1,\ldots,p_i-1\}$ — вычет по модулю числовой величины A по одному из модулей $p_i \in \{p_1,\ldots p_n\}$, принадлежащих полной упорядоченной системе оснований модулярной системы: $p_1 < p_2 \ldots < p_i < \ldots p_n$.

Векторы, определяемые соотношением (1), назовем модулярными.

Для анализа сходимости модулярных вычислительных процессов, в частности процессов, в которых входные, промежуточные и выходные числовые величины представлены в модулярной системе счисления (без выхода за ее пределы), рассмотрим свойства метрического векторного подпространства, элементами которого являются n-мерные модулярные векторы.

1. Свойства модулярного векторного подпространства

Введем n-мерное линейное векторное пространство V^n, элементы которого — векторы с компонентами, имеющими ограниченное значение. Для ограничения значения компонент используется операция вычисления наименьших неотрицательных или абсолютно наименьших вычетов по некоторому фиксированному модулю g, в частности по модулю максимального модулярного основания $g = p_n$:

$$A(\mathrm{mod}\,p_n^n) \leftrightarrow \\ \leftrightarrow (a_1(\mathrm{mod}\,p_n),\ldots,a_i(\mathrm{mod}\,p_n),\ldots a_n(\mathrm{mod}\,p_n)).$$

Компоненты таких векторов будем считать цифрами представления некоторой числовой величины A в позиционной системе с основанием p_n.

Векторное пространство V^n является линейным. Для его элементов выполняются условия однородности и аддитивности:

$$\forall k \in Z\ kA(\mathrm{mod}\,p_n^n) \leftrightarrow (ka_1(\mathrm{mod}\,p_n),\ldots, \\ \ldots, ka_i(\mathrm{mod}\,p_n),\ldots, ka_n(\mathrm{mod}\,p_n));$$

$$(A+B)\,(\mathrm{mod}\,p_n^n) \leftrightarrow ((a_1+b_1)(\mathrm{mod}\,p_n),\ldots, \\ \ldots,(a_i+b_i)(\mathrm{mod}\,p_n),\ldots.(a_n+b_n)(\mathrm{mod}\,p_n)).$$

В этом n-мерном векторном пространстве V^n введем подпространство W^n модулярных векторов, компоненты которых принадлежат полным системам вычетов по модулям соответствующих модулярных оснований. Векторное подпространство W^n назовем модулярным, его кольцевая структура при такой трактовке не учитывается. Элементами подпространства являются модулярные векторы, значения компонент которых ограничены результатами операций вычисления наименьших неотрицательных или абсолютно наименьших вычетов по отдельным модулям выбранной модулярной системы счисления. После определения скалярного произведения в векторном пространстве можно ввести ортогональное подпространство.

Модулярное подпространство W^n является линейным. Для его элементов

$$A(\mathrm{mod}\, P) \leftrightarrow$$
$$\leftrightarrow (a_1(\mathrm{mod}\, p_1),\ldots,a_i(\mathrm{mod}\, p_i),\ldots.a_n(\mathrm{mod}\, p_n));$$
$$B(\mathrm{mod}\, P) \leftrightarrow$$
$$\leftrightarrow (b_1(\mathrm{mod}\, p_1),\ldots,b_i(\mathrm{mod}\, p_i),\ldots.b_n(\mathrm{mod}\, p_n))$$

выполняются условия однородности и аддитивности:

$$\forall k \in Z\ kA(\mathrm{mod}\, P) \leftrightarrow (ka_1(\mathrm{mod}\, p_1),\ldots,$$
$$\ldots, ka_i(\mathrm{mod}\, p_i),\ldots,ka_n(\mathrm{mod}\, p_n));$$
$$(A+B)\ (\mathrm{mod}\, P) \leftrightarrow ((a_1+b_1)(\mathrm{mod}\, p_1),\ldots,$$
$$\ldots,(a_i+b_i)(\mathrm{mod}\, p_i),\ldots,(a_n+b_n)(\mathrm{mod}\, p_n)).$$

Операции над модулярными векторами — умножение на число и сложение — выполняются посредством вычислений вычетов компонент модулярных векторов по соответствующим модулям [1, 2].

2. Модулярные скалярные произведения

Для введения в подпространстве W^n модулярных векторов понятий ортогональности и базиса рассмотрим скалярные произведения.

Сформируем требования, которым должно удовлетворять модулярное скалярное произведение (A, B) двух векторов A, B, принадлежащих векторному модулярному подпространству W^n, учитывающее особенности модулярной системы счисления:

- симметричность, равенство $(A, B) = (B, A)$ должно выполняться для наименьших неотрицательных и абсолютно наименьших вычетов по модулю;
- умножение на число скалярного произведения $k(A, B)$ является произведением двух чисел;
- для выполнения аддитивности $((A + B)C) = (A, C) + (B, C)$ сумму векторов в модулярном пространстве определим следующим образом:

$$((A+B-kP),C) = (A,C)+(B,C)-k(P,C) =$$
$$= (A,C)+(B,C)+0;$$

- для выполнения условия неотрицательности модуля вектора используются наименьшие неотрицательные вычеты по модулю:

$$\forall A \in M\ (A,A) \geqslant 0,\ (A,A) = 0 \text{ при } A = \theta,$$

где θ — нулевой модулярный вектор.

Введем скалярные произведения векторов с модулярными компонентами, принадлежащих W^n.

Определение. В общем случае скалярное произведение следующего вида есть сумма скалярных произведений:

$$\left(\left(\sum_{i=1}^{k} A_i - kP\right), C\right) =$$
$$= \sum_{i=1}^{k}(A_i,C) + k(P,C) = \sum_{i=1}^{k}(A_i,C).$$

Определение. Скалярное произведение (первое) двух модулярных векторов A, $B \in W^n$ является числом:

$$(A,B) = \sum_{i=1}^{n} |a_i|_{p_i} |b_i|_{p_i}, \qquad (2)$$

где $|a_i|_{p_i}$ — обозначена бинарная операция вычисления вычета по модулю p_i в инфиксной записи.

Замечание $(A,B) < n(p_n - 1)^2$.

Введем модуль (первый) модулярного вектора как сумму квадратов вычетов по соответствующим модулям:

$$(A,A) = \sum_{i=1}^{n} |a_i|^2_{p_i},\ |A| = \sqrt{\sum_{i-1}^{n} |a_i|^2_{p_i}},$$
$$|A|^2 = \left(\sqrt{\sum_{i-1}^{n} |a_i|^2_{p_i}}\right)^2 = \sum_{i-1}^{n} |a_i|^2_{p_i}.$$

Замечание. В этих соотношениях и далее используется арифметическое значение квадратного корня.

Для вычисления скалярного произведения (2) возможно использование наименьших неотрицательных или абсолютно-наименьших вычетов по модулю.

Теорема Т-1. Для скалярного произведения (2) неравенство Коши—Буняковского имеет следующий вид:

$$\sum_{i-1}^{n}|a_i|_{p_i}|b_i|_{p_i} \leqslant \sqrt{\sum_{i-1}^{n}|a_i|_{p_i}^2}\sqrt{\sum_{i-1}^{n}|b_i|_{p_i}^2}.$$

Доказательство.

Сформируем выражение $C = |A|^2B - (A,B)A$, которое возведем в квадрат:

$$|C|^2 = (|A|^2B - (A,B)A)^2 =$$
$$= (|A|^2)^2|B|^2 - 2(A,B)^2|A|^2 + (A,B)^2|A|^2 =$$
$$= |A|^2(|A|^2|B|^2 - (A,B)^2).$$

Выполняется неравенство $|C|^2 \geqslant 0$. Следовательно, выполняются неравенства

$$(|A|^2|B|^2 - (A,B)^2) \geqslant 0,\ |A|^2|B|^2 \geqslant (A,B)^2 \text{ или}$$

$$|A||B| \geqslant (A,B) \text{ или } \sum_{i-1}^{n}|a_i|_{p_i}|b_i|_{p_i} \leqslant \sqrt{\sum_{i-1}^{n}|a_i|_{p_i}^2}\sqrt{\sum_{i-1}^{n}|b_i|_{p_i}^2}.$$

Неравенство выполняется при использовании наименьших неотрицательных вычетов или абсолютно наименьших вычетов по модулю.

Теорема Т-2. Для скалярного произведения (2) выполняется неравенство треугольника

$$\sqrt{\sum_{i=1}^{n}|(x_i - z_i)|_{p_i}^2} + \sqrt{\sum_{i=1}^{n}|(z_i - y_i)|_{p_i}^2} \geqslant \sqrt{\sum_{i=1}^{n}|(x_i - y_i)|_{p_i}^2}.$$

Доказательство.

Выполняются следующие модульные соотношения:

$$|x_i - y_i|_{p_i} = ||x_i - z_i|_{p_i} + |z_i - y_i|_{p_i}|_{p_i} =$$
$$= |x_i - z_i|_{p_i} + |z_i - y_i|_{p_i} - 0|p_i,$$

где символом $0|p_i$ обозначены альтернативные варианты "или".

Следовательно,

$$|x_i - z_i|_{p_i} + |z_i - y_i|_{p_i} \geqslant |x_i - y_i|_{p_i}.$$

Возведем обе части неравенства в квадрат и просуммируем:

$$\sum_{i=1}^{n}(|(x_i - z_i)|_{p_i}^2 + 2|(x_i - z_i)|_{p_i}|(z_i - y_i)|_{p_i} +$$
$$+ |(z_i - y_i)|_{p_i}^2) \geqslant \sum_{i=1}^{n}|(x_i - y_i)|_{p_i}^2$$

Применим неравенство Коши—Буняковского, которое для модульных соотношений в данном случае имеет вид

$$\sum_{i=1}^{n}|(x_i - z_i)|_{p_i}|(z_i - y_i)|_{p_i} \leqslant$$
$$\leqslant \sqrt{\sum_{i=1}^{n}|(x_i - z_i)|_{p_i}^2}\sqrt{\sum_{i=1}^{n}|(z_i - y_i)|_{p_i}^2}.$$

Получим

$$\left(\sqrt{\sum_{i=1}^{n}|(x_i - z_i)|_{p_i}^2}\right)^2 + 2\sqrt{\sum_{i=1}^{n}|(x_i - z_i)|_{p_i}^2}\sqrt{\sum_{i=1}^{n}|(z_i - y_i)|_{p_i}^2} +$$
$$+\left(\sqrt{\sum_{i=1}^{n}|(z_i - y_i)|_{p_i}^2}\right)^2 \geqslant \sum_{i=1}^{n}|(x_i - y_i)|_{p_i}^2$$

или

$$\left(\sqrt{\sum_{i=1}^{n}|(x_i - z_i)|_{p_i}^2} + \sqrt{\sum_{i=1}^{n}|(z_i - y_i)|_{p_i}^2}\right)^2 \geqslant \sum_{i=1}^{n}|(x_i - y_i)|_{p_i}^2$$

или

$$\sqrt{\sum_{i=1}^{n}|(x_i - z_i)|_{p_i}^2} + \sqrt{\sum_{i=1}^{n}|(z_i - y_i)|_{p_i}^2} \geqslant \sqrt{\sum_{i=1}^{n}|(x_i - y_i)|_{p_i}^2}.$$

Теорема доказана.

В символике векторной алгебры скалярное произведение (2) имеет вид

$$(A,B) = (a_1(\mathrm{mod}\ p_1),\ldots,a_n(\mathrm{mod}\ p_n))\times$$
$$\times(b_1(\mathrm{mod}\ p_1),\ldots,b_n(\mathrm{mod}\ p_n))^{\mathrm{T}},$$

где $^{\mathrm{T}}$ — символ транспонирования матрицы, вектора.

Введем модулярное (второе) скалярное произведение векторов $A,\ B \in W^n$.

Определение. Модулярное (второе) скалярное произведение двух модулярных векторов $A,\ B \in W^n$ есть число

$$(A,B)' = \sum_{i=1}^{n}|a_ib_i|_{p_i}. \qquad (3)$$

Замечание. $(A,B)' < n(p_n - 1)$.

Для модулярного скалярного произведения (3), учитывая взаимно однозначное соответствие, возможно использование наименьших неотрицательных и абсолютно наименьших вычетов по модулю.

Определим второй модуль модулярного вектора $A(\mathrm{mod}P)$ как сумму квадратичных вычетов по соответствующим модулям:

$$(A,A)' = \sum_{i=1}^{n}|a_ia_i|_{p_i} = \sum_{i-1}^{n}|a_i^2|_{p_i};$$

$$|A| = \sqrt{\sum_{i-1}^{n}|a_i^2|_{p_i}};\ |A|^2 = \left(\sqrt{\sum_{i-1}^{n}|a_i^2|_{p_i}}\right)^2 = \sum_{i-1}^{n}|a_i^2|_{p_i}.$$

Для вычисления модулей модулярных векторов используются наименьшие неотрица-

тельные вычеты, что не нарушает условие для скалярного произведения:

$$\forall A\ (A, A) \geqslant 0,\ (A, A) = 0 \text{ при } A = 0.$$

Теорема Т-3. Для модулярного скалярного произведения (3) неравенство Коши—Буняковского имеет следующий вид:

$$\sum_{i-1}^{n} |a_i b_i|_{p_i} \leqslant \sqrt{\sum_{i-1}^{n} |a_i^2|_{p_i}} \sqrt{\sum_{i-1}^{n} |b_i^2|_{p_i}}.$$

Доказательство.

Сформируем следующее выражение:

$$C = |A|^2 B - (A, B)' A.$$

Возведя выражение в квадрат, получим

$$|C|^2 = (|A|^2 B - (A, B)' A)^2 =$$
$$= (|A|^2)^2 |B|^2 - 2(A, B)'^2 |A|^2 + (A, B)'^2 |A|^2 =$$
$$= |A|^2 (|A|^2 |B|^2 - (A, B)'^2).$$

Учитывая, что $|C|^2 \geqslant 0$, можно заметить, что для второго числового сомножителя, записанного в виде разности квадрата модуля модулярного вектора и квадрата скалярного произведения (3), должно выполняться условие

$$(|A|^2 |B|^2 - (A, B)'^2) \geqslant 0,$$

следовательно,

$$|A|^2 |B|^2 \geqslant (A, B)'^2 \text{ и } (A, B)' \leqslant |A||B|$$
$$\text{или } \sum_{i-1}^{n} |a_i b_i|_{p_i} \leqslant \sqrt{\sum_{i-1}^{n} |a_i^2|_{p_i}} \sqrt{\sum_{i-1}^{n} |b_i^2|_{p_i}}.$$

Неравенство выполняется при использовании в левой части наименьших неотрицательных вычетов или абсолютно наименьших вычетов по модулю. В правой части неравенства возможно использование только наименьших неотрицательных вычетов по модулю, иначе нарушается соответствующее условие для скалярного произведения.

Теорема Т-4. Для модулярного скалярного произведения (3) выполняется неравенство треугольника

$$\sqrt{\sum_{i=1}^{n} |(x_i - z_i)^2|_{p_i}} + \sqrt{\sum_{i=1}^{n} |(z_i - y_i)^2|_{p_i}} \geqslant$$
$$\geqslant \sqrt{\sum_{i=1}^{n} |(x_i - y_i)^2|_{p_i}}.$$

Доказательство.

Выполняются следующие модульные соотношения:

$$|x_i - y_i|_{p_i} = \| x_i - z_i |_{p_i} + | z_i - y_i |_{p_i} |_{p_i} =$$
$$= |x_i - z_i|_{p_i} + |z_i - y_i|_{p_i} - 0|p_i,$$

где символом $0|p_i$ обозначены альтернативные варианты "или". Следовательно,

$$|x_i - z_i|_{p_i} + |z_i - y_i|_{p_i} \geqslant |x_i - y_i|_{p_i}.$$

Возведем в квадрат обе части неравенства

$$(|x_i - z_i|_{p_i} + |z_i - y_i|_{p_i})^2 \geqslant (|x_i - y_i|_{p_i})^2,$$

получим

$$|x_i - z_i|_{p_i}^2 + |x_i - z_i|_{p_i} |z_i - y_i|_{p_i} + |x_i - z_i|_{p_i} |z_i - y_i|_{p_i} + \\ + |z_i - y_i|_{p_i}^2 \geqslant (|x_i - y_i|_{p_i})^2. \quad (4)$$

Возьмем вычеты по модулю от каждого слагаемого в обеих частях неравенства (4) и выполним преобразования по модулю в левой части неравенства:

$$\| (x_i - z_i)^2 |_{p_i} + |(x_i - z_i)(z_i - y_i)|_{p_i} +$$
$$+ |(x_i - z_i)(z_i - y_i)|_{p_i} + |(z_i - y_i)^2|_{p_i} |_{p_i} \geqslant$$
$$\geqslant \left| (|x_i - y_i|_{p_i})^2 \right|_{p_i};$$
$$\| (x_i - z_i)(x_i - z_i + z_i - y_i) |_{p_i} +$$
$$+ |(x_i - z_i + z_i - y_i)(z_i - y_i)|_{p_i} |_{p_i} \geqslant$$
$$\geqslant \left| (|x_i - y_i|_{p_i})^2 \right|_{p_i}.$$

В результате получим соотношение, которое показывает, что левая часть не меньше правой части:

$$\| (x_i - y_i) |_{p_i} (x_i - y_i) |_{p_i} |_{p_i} \geqslant \left| (x_i - y_i)^2 \right|_{p_i}.$$

Заметим, что при $z = 0$ нестрогое неравенство превращается в равенство.

После возведения в квадрат обеих частей неравенства (4) и вычисления вычетов по модулю окончательно получим слева сумму четырех вычетов по модулю, которая не меньше правой части, содержащей одиночный вычет по модулю:

$$|(x_i - z_i)^2|_{p_i} + |(x_i - z_i)(z_i - y_i)|_{p_i} +$$
$$+ |(x_i - z_i)(z_i - y_i)|_{p_i} + |(z_i - y_i)^2|_{p_i} \geqslant \left| (x_i - y_i)^2 \right|_{p_i}.$$

После суммирования в обеих частях неравенства получим

$$\sum_{i=1}^{n} (|(x_i - z_i)^2|_{p_i} + 2|(x_i - z_i)(z_i - y_i)|_{p_i} +$$
$$+ |(z_i - y_i)^2|_{p_i}) \geqslant \sum_{i=1}^{n} |(x_i - y_i)^2|_{p_i}.$$

Применим неравенство Коши—Буняковского, которое для данного случая имеет вид

$$\sum_{i=1}^{n}|(x_i - z_i)(z_i - y_i)|_{p_i} \leqslant$$
$$\leqslant \sqrt{\sum_{i=1}^{n}|(x_i - z_i)^2|_{p_i}}\sqrt{\sum_{i=1}^{n}|(z_i - y_i)^2|_{p_i}}.$$

В результате получим:

$$\left(\sqrt{\sum_{i=1}^{n}|(x_i - z_i)^2|_{p_i}}\right)^2 + 2\sqrt{\sum_{i=1}^{n}|(x_i - z_i)^2|_{p_i}}\times$$
$$\times\sqrt{\sum_{i=1}^{n}|(z_i - y_i)^2|_{p_i}} + \left(\sqrt{\sum_{i=1}^{n}|(z_i - y_i)^2|_{p_i}}\right)^2 \geqslant$$
$$\geqslant \sum_{i=1}^{n}|(x_i - y_i)^2|_{p_i}$$

или

$$\left(\sqrt{\sum_{i=1}^{n}|(x_i - z_i)^2|_{p_i}} + \sqrt{\sum_{i=1}^{n}|(z_i - y_i)^2|_{p_i}}\right)^2 \geqslant$$
$$\geqslant \sum_{i=1}^{n}|(x_i - y_i)^2|_{p_i}$$

или

$$\sqrt{\sum_{i=1}^{n}|(x_i - z_i)^2|_{p_i}} + \sqrt{\sum_{i=1}^{n}|(z_i - y_i)^2|_{p_i}} \geqslant$$
$$\geqslant \sqrt{\sum_{i=1}^{n}|(x_i - y_i)^2|_{p_i}}.$$

Теорема доказана.

Определение. Модулярный оператор линейной свертки $S(A \cdot B(\mathrm{mod}\, P))$ двух модулярных векторов $A,\ B \in W^n$ вычисляется следующим образом:

$$S(A \cdot B(\mathrm{mod}\, P)) \leftrightarrow$$
$$\leftrightarrow \begin{pmatrix} a_1 \ldots & a_i \ldots & a_n \\ . & . & . \\ a_1 \ldots & a_i \ldots & a_n \ldots \end{pmatrix} \cdot (b_1, \ldots, b_i, \ldots, b_n)^{\mathrm{T}} = \qquad (5)$$
$$= \left(\sum_{i=1}^{n}|a_i b_i|_{p_i}(\mathrm{mod}\, p_1), \ldots, \sum_{i=1}^{n}|a_i b_i|_{p_i}(\mathrm{mod}\, p_n)\right)^{\mathrm{T}}.$$

Рассмотрим вопрос базиса в модулярном векторном подпространстве W^n. В векторном модулярном подпространстве W^n существует как минимум один ортонормированный базис, состоящий из базисных векторов модулярной системы счисления:

$$\forall i = 1, \ldots, n\ B_i(\mathrm{mod}\, P) \leftrightarrow (0(\mathrm{mod}\, p_1), \ldots,$$
$$\ldots, m_i \frac{P}{p_i}(\mathrm{mod}\, p_i), \ldots, 0(\mathrm{mod}\, p_n)) =$$
$$= (0(\mathrm{mod}\, p_1), \ldots, 1(\mathrm{mod}\, p_i), \ldots, 0(\mathrm{mod}\, p_n)),$$

где $\forall i = 1, \ldots, n\ \ m_i = \left|\frac{P}{p_i}\right|_{p_i}^{-1}$.

Базис ортонормированный, скалярные произведения (2) и (3) взятых попарно базисных векторов равны нулю [3, 4]:

$$(B_i, B_j) = (B_i, B_j)' =$$
$$= (0, \ldots, 1(\mathrm{mod}\, p_i), \ldots, 0) \cdot (0, \ldots, 1(\mathrm{mod}\, p_j), \ldots, 0)^{\mathrm{T}} = 0.$$

Базисом в векторном пространстве является любая совокупность линейно независимых векторов, он может быть найден стандартными методами. Решение неоднородной системы линейных алгебраических уравнений, столбцами которой являются базисные векторы, дает разложение по этому базису произвольного вектора, записываемого в правой части системы уравнений.

3. Метрики в векторном модулярном подпространстве

Для анализа сходимости модулярных вычислительных процессов введем два вида расстояний или метрик (аналогов евклидового расстояния) на множестве n-мерных модулярных векторов из подпространства W^n через их скалярные произведения.

Определение. Модулярное расстояние между двумя модулярными величинами — векторами $A,\ B \in W^n$ — на основе скалярного произведения (2) есть число

$$l^1(A, B) = \sqrt{\sum_{i=1}^{n}|a_i - b_i|_{p_i}^2}.$$

Определение. Модулярный вес модулярного вектора A на основе скалярного произведения (2) есть модулярное расстояние между произвольным модулярным вектором и нулевым:

$$w^1(A) = l^1(A, \theta) = \sqrt{\sum_{i=1}^{n}|a_i|_{p_i}^2} = \sqrt{(A, A)}.$$

Модулярный вес, полученный на основе скалярного произведения (2), совпадает с первым модулем вектора в модулярном векторном подпространстве W^n.

Теорема Т-5. Модулярное расстояние $l^1(A, B)$ между двумя модулярными векторами

$A, B \in W^n$ на основе скалярного произведения (2) является метрикой.

Доказательство.

Используя арифметическое значение квадратного корня, можно получить:

1. $l^1(A, B) \geqslant 0$.
2. $l^1(A, B) = 0$, если $A \equiv B(\mathrm{mod}\, P)$.
3. $l^1(A, B) = l^1(B, A)$.
4. Выполнение неравенства треугольника установлено в теореме *Т-2*.

Определение. Модулярное расстояние между двумя модулярными величинами — векторами $A, B \in W^n$ — на основе модулярного скалярного произведения (3) есть число

$$l^2(A, B) = \sqrt{\sum_{i=1}^{n} |(a_i - b_i)^2|_{p_i}}.$$

Определение. Модулярный вес модулярного вектора A на основе модулярного скалярного произведения (3) есть модулярное расстояние между произвольным модулярным вектором и нулевым:

$$w^2(A) = l^2(A, \theta) = \sqrt{\sum_{i=1}^{n} |a_i^2|_{p_i}} = \sqrt{(A, A)'}.$$

Модулярный вес, полученный на основе модулярного скалярного произведения (3), совпадает со вторым модулем вектора в модулярном векторном подпространстве W^n.

Модулярные веса, полученные на основе скалярных произведений (2) и (3), не равны в общем случае.

Теорема Т-6. Модулярное расстояние $l^2(A, B)$ между двумя модулярными векторами $A, B \in W^n$ на основе модулярного скалярного произведения (3) является метрикой.

Доказательство.

Используя арифметическое значение квадратного корня, можно получить:

1. $l^2(A, B) \geqslant 0$.
2. $l^2(A, B) = 0$, если $A \equiv B(\mathrm{mod}\, P)$.
3. $l^2(A, B) = l^2(B, A)$.
4. Выполнение неравенства треугольника установлено в теореме *Т-4*.

Приведем для полноты обзора, кроме выше введенных метрик, остаточное расстояние (аналог метрики Хэмминга), применимое для задач помехозащитного модулярного кодирования дискретной информации и учитывающее характер модульной ошибки канала передачи, хранения и обработки для векторов с модулярными компонентами [5,6].

Определение. Остаточное расстояние $d(A, B)$ между двумя модулярными векторами $A, B \in W^n$ есть остаточный вес модульной разности двух модулярных векторов

$$d(A, B) = \sum_{i=1}^{n} \delta(|a_i - b_i|_{p_i}),$$

где символ Кронекера

$$\delta(|a_i - b_i|_{p_i}) = \begin{cases} 1, \text{ при } a_i \neq b_i; \\ 0, \text{ при } a_i = b_i. \end{cases}$$

Остаточное расстояние является метрикой, так как для него выполняются соответствующие аксиомы [6].

Остаточный вес определяется как остаточное расстояние между произвольным модулярным вектором и нулевым.

Введем модулярный аналог расстояния и веса Ли, предназначенный для помехозащитного кодирования дискретной информации в системах передачи данных с фазовой модуляцией несущего сигнала.

Определение. Модулярный вес Ли $|a_i|_L$ одиночной компоненты a_i модулярного вектора равен:

$$|a_i|_L = a_i \text{ при } 0 \leqslant a_i \leqslant \frac{p_i - 1}{2};$$

$$|a_i|_L = p_i - a_i \text{ при } \frac{p_i - 1}{2} < a_i \leqslant p_i - 1.$$

Определение. Модулярное расстояние Ли между двумя модулярными векторами $A, B \in W^n$ есть сумма модулярных весов Ли разности их компонент:

$$t(A, B) = \sum_{i=1}^{n} ||a_i - b_i|_{p_i}|_L.$$

В вычислительных экспериментах для ускорения вычислений нормированных компонент модулярных векторов, а также ряда функций, например оператора линейной свертки, необходимые для этих процедур константы хранятся в специальных областях кэш-памяти моделируемой модулярной вычислительной системы [7].

Заключение

Известны классы вычислительных процессов, оперирующих с числовыми величинами и называемых многоразрядными процессами или процессами в больших компьютерных диапазонах. В этих процессах операнды, промежуточные результаты операций и результаты вычислительных процессов являются модулярными числовыми величинами, т.е. представленными в компьютерной моду-

лярной системе счисления [3, 8]. Модулярные представления для целых числовых величин и правильных рациональных дробей при соответствующих алгоритмах позволяют организовать эффективное распараллеливание вычислительного процесса [8, 9]. Технической базой таких параллельных процессов являются вычислительные системы с SIMD-архитектурой, лежащей в основе большинства современных многопроцессорных систем, содержащих кроме центральных процессоров CPU множество графических ускорителей GPU, используемых для распараллеливания вычислений [2, 9].

По оценкам ряда исследователей модулярная машинная арифметика имеет преимущества именно в области параллельных многоразрядных вычислений, она позволяет организовывать высокопроизводительные модулярные вычислительные процессы в больших компьютерных диапазонах [9, 10].

Анализ особенностей модулярного векторного подпространства, введенные модулярные скалярные произведения позволяют определить метрики в векторном модулярном пространстве W^n, предназначенные для оценки сходимости модулярных вычислительных процессов.

Список литературы

1. **Амербаев В. М.** Теоретические основы машинной арифметики. Алма-Ата: Наука, 1976. 320 с.
2. **Ахо А.** и др. Построение и анализ вычислительных алгоритмов. М.: Мир, 2011. 536 с.
3. **Инютин С. А.** Основы модулярной алгоритмики. Ханты-Мансийск: Полиграфист, 2009. 237 с.
4. **Inutin S.** Parallel Square Modular Computer Algebra. Transaction of Parallel Processing and Applied Mathematics. Poland-Denmark: Springer, LNCS 3019, 2003. P. 539—547.
5. **Инютин С. А**. Проблема метрик в модулярном помехозащитном кодировании // Труды СурГУ. 2008. Вып. 12. С. 84—93.
6. **Торгашев В. А.** Система остаточных классов и надежность ЦВМ. М.: Советское радио, 1973. 120 с.
7. **Столярский Е. З., Шилов В. В.** Организация и работа кэш-памяти // Информационные технологии. 2000. № 7. С. 2—8.
8. **Инютин С. А.** Анализ сложности многоразрядных вычислительных процессов // Научные труды МАТИ. 2014. Вып. 22 (94). С. 154—159.
9. **Инютин С. А**. Комплексирование систем счисления для многоразрядных вычислительных процессов // Информационные технологии.2018. Т. 24, № 12. С. 343—347.
10. **Ноден П., Китте К.** Алгебраическая алгоритмика. М.: Мир, 1999. 720с.

S. A. Inyutin, Doctor Technical Science (PhD), Full Professor, e-mail: inyutin_int@mail.ru,
Moscow Aviation Institute (Nation Research University) (MAI), Moscow, Russian Federation

Metrics for Modular Vectors Space

Modular representations of numerical quantities in the form of residue vectors by prime modules from a finite set can be considered as a set of modular vectors. The set of modular vectors is considered as a linear subspace in a vector space containing vectors with components of limited size; its ring structure is not taken into account. The properties of a linear modular subspace, methods for determining new scalar products are analyzed, which allows us to introduce new metrics. The introduced algebraic constructions are designed to analyze the convergence of multi-bit parallel computing processes in large computer ranges that operate with numerical values in the modular representation. The introduced numerical vector representations are oriented for application in modular reconfigurable computing systems of SIMD architecture.

Keywords: *modular vector subspace, modular scalar products, modular metrics, parallel modular computing process, multiprocessor reconfigurable system*

DOI: 10.17587/it.26.570-576

References

1. **Amerbaev V. M.** Theoretic base computer arithmetic, Alma-Ata, Nauka, 1976, 320 p.
2. **Aho A., Hopcroft J., Ullman J.** The design and analysis of computer algorithms, Moscow, Mir, 2011, 536 p.
3. **Inyutin S. A.** Base at modular algorithmic, Hanty-Mansiysk, Poligrafist, 2009, 237 p.
4. **Inutin S.** Parallel Square Modular Computer Algebra. Transaction of Parallel Processing and Applied Mathematics, Poland-Denmark, Springer, LNCS 3019, 2003, pp. 539—547.
5. **Inyutin S. A.** The problem of metrics in modular error control-codes, *Transactions of SUSU*, vol. 12, Surgut, RIO, 2008, pp. 84—93.
6. **Torgashev V. A.** The system of residual classes and the reliability of the computer, Moscow, Soviet Radio, 1973, 120 p.
7. **Stolyarskiy E. Z., Shilov V. V.** Cache organization and operation, *Informatsionnyie Tehnologii*, 2000, no. 7, pp. 2—8.
8. **Inyutin S. A.** Analysis of many digital calculation process, *Nauchnye trudy MATI,* 2014, vol. 22 (94), pp. 154—159.
9. **Inyutin S. A.** Integration of number systems for multi-digit computing processes, *Informatsionnyie Tehnologii*, 2018, no. 12, vol. 26, pp. 343—347.
10. **Noden P., Kitte K.** Algebraic algorithmic, Moscow, Mir, 1999, 720 p.

УДК 003.26 DOI: 10.17587/it.26.577-585

Р. Ш. Фахрутдинов, канд. техн. наук, зав. лабораторией, e-mail: fahr@cobra.ru,
А. Ю. Мирин, канд. техн. наук, ст. науч. сотр., e-mail: mirin@cobra.ru,
Д. Н. Молдовян, канд. техн. наук, науч. сотр., e-mail: mdn.spectr@mail.ru,
А. А. Костина, науч. сотр., e-mail: anna-kostina1805@mail.ru,
Санкт-Петербургский институт информатики и автоматизации Российской академии наук,
г. Санкт-Петербург

Схемы открытого согласования ключей на основе скрытой задачи дискретного логарифмирования[1]

Рассмотрены схемы открытого согласования ключа, основанные на вычислительной сложности скрытой задачи дискретного логарифмирования, задаваемой в конечных некоммутативных ассоциативных алгебрах. Для повышения производительности предложены алгебры с заданием операции векторного умножения с помощью прореженных таблиц умножения базисных векторов и процедура генерации перестановочных ключевых элементов, свободная от операций экспоненцирования.

Ключевые слова: *защита информации, криптография, открытое согласование ключей, задача дискретного логарифмирования, конечная ассоциативная алгебра, некоммутативная алгебра, глобальная единица, локальная единица, левосторонняя единица*

Введение

В настоящее время прогресс в области квантовых вычислений достиг такого уровня, при котором высокую степень актуальности приобрела проблема разработки криптографических алгоритмов и протоколов с открытым ключом, обеспечивающих высокую стойкость к квантовым атакам, т.е. атакам с использованием квантового компьютера [1—3]. Криптосхемы, удовлетворяющие этому условию, называются постквантовыми, т.е. ориентированными на применение в эпоху квантовых вычислений, наступление которой прогнозируется на ближайшее будущее [4].

Указанная проблема связана с тем, что современные двухключевые криптосхемы, имеющие широкое применение для решения многочисленных задач в области обеспечения информационной безопасности и кибербезопасности, основаны на вычислительной трудности задачи дискретного логарифмирования (ЗДЛ) и задачи факторизации (ЗФ), которые на квантовом компьютере могут быть решены за полиномиальное время [5—8]. Полиномиальные квантовые алгоритмы решения ЗДЛ и ЗФ используют возможность сведения этих задач к задаче определения длины периода некоторой периодической функции, принимающей дискретные значения в рамках явно заданной конечной циклической группы. В частности, при решении ЗДЛ строится периодическая функция, длина одного из периодов которой зависит от искомого значения логарифма. Квантовый компьютер чрезвычайно эффективно реализует дискретное преобразование Фурье, из максимумов которого вычисляются длины периодов преобразованной функции [9, 10].

Для разработки постквантовых криптографических алгоритмов и протоколов используют вычислительно трудные задачи, отличные от ЗДЛ и ЗФ. Перспективным подходом к разработке постквантовых двухключевых криптосхем является применение в качестве базового криптографического примитива скрытой задачи дискретного логарифмирования (СЗДЛ) [11—13]. На основе этого примитива предложены протоколы открытого согласования ключа [14] и электронной цифровой подписи (ЭЦП) [15, 16], а также алгоритмы коммутативного

[1] Работа выполнена при поддержке бюджетной темы № 0060-2019-0010.

шифрования [17, 18]. В качестве алгебраического носителя таких криптосхем используются некоммутативные конечные ассоциативные алгебры (НКАА) [19—22], содержащие в себе большое число различных конечных циклических групп в качестве подмножеств элементов алгебры.

В данной статье рассматриваются известные протоколы открытого согласования ключа, построенные на основе вычислительной трудности СЗДЛ, и предлагается и обосновывается новый способ построения криптосхем данного типа, который позволяет повысить их вычислительную эффективность. Предлагаемый способ отличается использованием в качестве алгебраического носителя криптосхем НКАА с операцией умножения, задаваемой по прореженным таблицам умножения базисных векторов (ТУБВ).

1. Задание скрытой задачи дискретного логарифмирования

Традиционная ЗДЛ формулируется следующим образом. Известен открытый ключ Y', представляющий собой некоторый элемент конечной циклической группы, вычисленный путем выполнения операции возведения в целочисленную степень достаточно большой разрядности:

$$Y' = N^x,$$

где N — генератор конечной циклической группы; x — секретный ключ. Нахождение значения x по известным элементам N и Y' называется ЗДЛ. В случае группы, имеющей значение ее порядка, равное многоразрядному простому числу q (длиной 256 бит и более), для обычного компьютера известны только сверхполиномиальные алгоритмы решения ЗДЛ, задаваемой в мультипликативной группе поля $GF(p)$ и в ряде других конечных групп.

На квантовом компьютере ЗДЛ решается как задача вычисления длины периода функции $f(i, j) = Y'^i N^j$ с натуральными значениями i и j, которая содержит период длины $(-1, x)$:

$$Y'^i N^j = Y'^{i-1} N^{j+x} \Rightarrow f(i, j) = f(i-1, j+x).$$

Для функции $f(i, j)$ со значениями в явно заданной конечной циклической группе квантовый компьютер эффективно выполняет дискретное преобразование Фурье, что позволяет за полиномиальное время найти длины всех периодов функции $f(i, j)$, в том числе и значение $(-1, x)$, а следовательно, и значение дискретного логарифма x.

Различные формы СЗДЛ возникают при построении двухключевых криптосхем, в которых основной операцией, определяющей высокий уровень стойкости, является операция возведения в степень, выполняемая в некоторой скрытой циклической группе, содержащейся в НКАА. Маскирование этой группы реализуется тем, что оба элемента N и N^x группы или один из них предоставляются в виде открытых параметров Y и Z криптосхемы после дополнительного маскирующего преобразования с помощью операций ψ_1 и ψ_2, являющихся взаимно коммутативными с операцией возведения в степень экспоненцирования: $Y = \psi_1(N^x)$ и $Z = \psi_2(N)$. При этом секретные маскирующие операции выбираются такими, что значения Y и Z лежат в разных циклических группах, каждая из которых отлична от группы, генерируемой всевозможными степенями элемента N.

Для корректности работы схемы ЭЦП маскирующие операции ψ_1 и ψ_2 должны быть согласованы между собой, что дает возможность построения периодических функций по значениям элементов открытого ключа. Например, в схемах ЭЦП, описанных в работах [23—25], открытым ключом является пара значений Y и Z, по которым может быть построена периодическая функция $f(i, j) = Y^i Z$, включающая период, имеющий длину $(-1, x)$, однако эта функция принимает значения, лежащие во многих различных циклических группах. Это обеспечивает стойкость к квантовым атакам на основе известных квантовых алгоритмов вычисления длины периода.

В схемах открытого согласования ключей предполагается, что независимые пользователи выполняют базовую операцию экспоненцирования в одной и той же циклической группе, поэтому кроме открытого ключа в криптосхемах данного типа должен быть задан другой известный параметр, позволяющий выполнить указанное условие. В рассматриваемом случае маскирование базовой циклической группы обеспечивается только маскированием результата выполнения операции экспоненцирования, т.е. открытый ключ включает элементы $Y = \psi_1(N^x)$ и N. Однако, несмотря на кажущуюся ослабленной маскировку, периодические функции, построенные с использованием значений Y и N, не содержат период, определяемый значением дискретного логарифма, а включают периоды, связанные со значением порядка q базовой циклической группы.

2. Некоммутативные конечные алгебры с ассоциативной операцией умножения

Конечные m-мерные алгебры представляют собой конечные m-мерные векторные пространства, заданные над конечным полем, например, над простым конечным полем $GF(p)$, в которых дополнительно к имеющимся операциям сложения векторов и скалярного умножения определена операция векторного умножения (далее операция умножения), которая является дистрибутивной слева и справа относительно операции сложения. Для построения двухключевых криптосхем на основе СЗДЛ используются НКАА, поэтому при задании операции умножения в векторном пространстве применяется способ определения этой операции по таблицам умножения базисных векторов (ТУБВ), позволяющий обеспечить свойство ассоциативности операции умножения.

Элементами m-мерного векторного пространства являются всевозможные векторы вида

$$\mathbf{A} = (a_0, a_1, \ldots, a_{m-1}) = a_0\mathbf{e}_0 + a_1\mathbf{e}_1 + \ldots + a_{m-1}\mathbf{e}_{m-1},$$

где $a_i \in GF(p)$, где p — простое число; $\mathbf{e}_i$ — формальные базисные векторы. Операция умножения ($\circ$) векторов $\mathbf{A} = \sum_{i=0}^{m-1} a_i\mathbf{e}_i$ и $\mathbf{B} = \sum_{j=0}^{m-1} b_j\mathbf{e}_j$ определяется по следующей формуле

$$\mathbf{A} \circ \mathbf{B} = \sum_{i=0}^{m-1}\sum_{j=0}^{m-1} a_i b_j \mathbf{e}_i \circ \mathbf{e}_j, \qquad (1)$$

где каждое из произведений пар базисных векторов заменяется на однокомпонентный вектор $\lambda\mathbf{e}_k$, задаваемый специально разработанной ТУБВ. Значение $\lambda \in GF(p)$ называется структурной константой. Рассмотрим произведение трех векторов $\mathbf{A}$, $\mathbf{B}$ и $\mathbf{C} = \sum_{k=0}^{m-1} c_k\mathbf{e}_k$, осуществляемое в соответствии со следующими двумя вариантами:

$$(\mathbf{A} \circ \mathbf{B}) \circ \mathbf{C} = \left(\sum_{i=0}^{m-1}\sum_{j=0}^{m-1} a_i b_j \mathbf{e}_i \circ \mathbf{e}_j\right) \circ \sum_{k=0}^{m-1} c_k\mathbf{e}_k = \sum_{i=0}^{m-1}\sum_{j=0}^{m-1}\sum_{k=0}^{m-1} a_i b_j c_k (\mathbf{e}_i \circ \mathbf{e}_j) \circ \mathbf{e}_k; \qquad (2)$$

$$\mathbf{A} \circ (\mathbf{B} \circ \mathbf{C}) = \left(\sum_{i=0}^{m-1} a_i\mathbf{e}_i\right) \circ \left(\sum_{j=0}^{m-1}\sum_{k=0}^{m-1} b_j c_k \mathbf{e}_i \circ \mathbf{e}_j\right) = \sum_{i=0}^{m-1}\sum_{j=0}^{m-1}\sum_{k=0}^{m-1} a_i b_j c_k \mathbf{e}_i \circ (\mathbf{e}_j \circ \mathbf{e}_k). \qquad (3)$$

Таблица 1

Задание операции умножения четырехмерной НКАА ($\mu \neq 0$; $\lambda \neq 0$)

$\circ$	$\mathbf{e}_0$	$\mathbf{e}_1$	$\mathbf{e}_2$	$\mathbf{e}_3$
$\mathbf{e}_0$	$\mu\mathbf{e}_0$	0	0	$\mu\mathbf{e}_3$
$\mathbf{e}_1$	0	$\lambda\mathbf{e}_1$	$\lambda\mathbf{e}_2$	0
$\mathbf{e}_2$	$\mu\mathbf{e}_2$	0	0	$\mu\mathbf{e}_1$
$\mathbf{e}_3$	0	$\lambda\mathbf{e}_3$	$\lambda\mathbf{e}_0$	0

Таблица 2

Задание операции умножения в шестимерной НКАА ($\lambda \neq 0$)

$\circ$	$\mathbf{e}_0$	$\mathbf{e}_1$	$\mathbf{e}_2$	$\mathbf{e}_3$	$\mathbf{e}_4$	$\mathbf{e}_5$
$\mathbf{e}_0$	$\mathbf{e}_0$	0	$\mathbf{e}_2$	0	$\mathbf{e}_4$	0
$\mathbf{e}_1$	$\lambda\mathbf{e}_3$	0	$\lambda\mathbf{e}_5$	0	$\lambda\mathbf{e}_1$	0
$\mathbf{e}_2$	0	$\mathbf{e}_4$	0	$\mathbf{e}_0$	0	$\mathbf{e}_2$
$\mathbf{e}_3$	$\mathbf{e}_3$	0	$\mathbf{e}_5$	0	$\mathbf{e}_1$	0
$\mathbf{e}_4$	$\lambda\mathbf{e}_0$	0	$\lambda\mathbf{e}_2$	0	$\lambda\mathbf{e}_4$	0
$\mathbf{e}_5$	0	$\mathbf{e}_1$	0	$\mathbf{e}_3$	0	$\mathbf{e}_5$

Равенство правых частей выражений (2) и (3) имеет место, если используемая ТУБВ обеспечивает выполнение равенства

$$(\mathbf{e}_i \circ \mathbf{e}_j) \circ \mathbf{e}_k = \mathbf{e}_i \circ (\mathbf{e}_j \circ \mathbf{e}_k) \qquad (4)$$

для всех возможных троек значений (i, j, k).

Известен ряд частных ТУБВ для реализации НКАА с фиксированным значением размерности векторов [19] и унифицированные способы построения ТУБВ [20—22] для произвольных четных значений размерности m. Из формулы (1) легко видеть, что вычислительная сложность операции умножения определяется вычислительной сложностью операции умножения в поле $GF(p)$ и числом этих операций, поэтому для уменьшения сложности базовой операции экспоненцирования (т.е. для повышения производительности разрабатываемых криптосхем) представляет интерес разработка прореженных ТУБВ, в которых в достаточно большом числе ячеек содержится структурный коэффициент, равный нулю.

Для случаев размерностей $m = 4$ и $m = 6$ в данном исследовании разработаны и используются ТУБВ указанного типа, которые показаны как табл. 1 и табл. 2.

3. Свойства четырехмерной НКАА

Решение векторных уравнений вида

$$\mathbf{X} \circ \mathbf{A} = \mathbf{A} \qquad (5)$$

и

$$\mathbf{A} \circ \mathbf{X} = \mathbf{A}, \qquad (6)$$

где $\mathbf{A} = (a_0, a_1, a_2, a_3)$ — некоторый заданный четырехмерный вектор; $\mathbf{X} = (x_0, x_1, x_2, x_3)$ — неизвестный вектор, приводит к получению следующей формулы для глобальной двухсторонней единицы **E**, содержащейся в алгебре:

$$\mathbf{E} = (\mu^{-1}, \lambda^{-1}, 0, 0). \quad (7)$$

Здесь используется термин "глобальная" для обозначения того, что данная единица действует на все элементы алгебры как двухсторонняя единица (в отличии от локальных единиц, действующих в подмножествах элементов алгебры).

Для преобладающего множества четырехмерных векторов рассматриваемой алгебры, координаты которых удовлетворяют условию $a_0a_1 \neq a_2a_3$, уравнения (5) и (6) имеют единственное решение $\mathbf{X} = \mathbf{E}$ (в случае $a_0a_1 = a_2a_3$ кроме этого решения имеется много других решений). Векторы, удовлетворяющие условию $a_0a_1 \neq a_2a_3$, являются обратимыми, т. е. для них каждое из векторных уравнений $\mathbf{X} \circ \mathbf{A} = \mathbf{A}$ и $\mathbf{A} \circ \mathbf{X} = \mathbf{A}$, имеет единственное решение $\mathbf{X} = \mathbf{A}^{-1}$, которое называется обратным значением вектора **A**. В случае $a_0a_1 = a_2a_3$ последние два векторных уравнения не имеют решений. Векторы, удовлетворяющие последнему условию, называются необратимыми. Из условия необратимости легко установить число необратимых векторов, которое равно $p^3 + p^2 - p$. Число всех четырехмерных векторов равно значению p^4, поэтому для числа обратимых векторов Ω (порядок мультипликативной группы алгебры) получаем следующую формулу:

$$\Omega = p(p-1)(p^2-1). \quad (8)$$

Каждый обратимый вектор $\mathbf{V} = (v_0, v_1, v_2, v_3)$ задает операцию автоморфного отображения, описываемую формулой

$$\varphi(\mathbf{X}) = \mathbf{V} \circ \mathbf{X} \circ \mathbf{V}^{-1}, \quad (9)$$

где переменная **X** пробегает все значения алгебры, которая является взаимно коммутативной с операцией экспоненцирования и представляет интерес как маскирующая операция при задании СЗДЛ.

В случае построения схем открытого согласования ключей маскирующие операции выполняются двумя пользователями в различной очередности, и порядок их выполнения не должен влиять на значение получаемого результата (общего секретного ключа), поэтому выбираемые пользователями маскирующие операции должны быть взаимно коммутативными, оставаясь при этом секретными. Взаимная коммутативность маскирующих операций, выбираемых любыми двумя пользователями, может быть обеспечена, если задать выбор вектора **V** как параметра операции автоморфного отображения из одной и той же коммутативной группы, содержащейся в алгебре. При этом порядок этой коммутативной группы должен быть достаточно большим. В работах [23, 24] эта проблема решается путем задания в качестве открытого параметра криптосхемы некоторого вектора **Q**, имеющего достаточно большой порядок ω, и механизма выбора вектора **V** путем использования случайного натурального числа $x < \omega$ и вычисления значения $\mathbf{V} = \mathbf{Q}^x$.

Недостатками этого способа являются ограничение множества потенциально возможных значений **V** и необходимость выполнения операции экспоненцирования, что повышает вычислительную трудоемкость процедуры формирования ключевых параметров пользователей. Для устранения этих недостатков предлагается способ задания полной коммутативной группы для выбора параметров маскирующей операции автоморфного отображения алгебры. Способ описывается следующим образом:

1) выбирается некоторый вектор

$$\mathbf{Q} = (q_0, q_1, q_2, q_3);$$

2) выводится формула, описывающая все векторы, которые перестановочны с **Q**;

3) по указанной формуле каждый пользователь вычисляет случайный вектор **V**, используемый для задания секретной операции маскирования.

Для векторов **X**, являющихся перестановочными с **Q**, выполняется условие $\mathbf{X} \circ \mathbf{Q} = \mathbf{Q} \circ \mathbf{X}$, т. е. множество векторов **X** являются решениями векторного уравнения $\mathbf{X} \circ \mathbf{Q} - \mathbf{Q} \circ \mathbf{X} = (0,0,0,0)$, которое сводится к решению следующей системы из четырех линейных уравнений с неизвестными x_0, x_1, x_2 и x_3:

$$\begin{cases} \mu x_0 q_0 + \lambda x_3 q_2 - \mu x_0 q_0 - \lambda x_2 q_3 = 0; \\ \lambda x_1 q_1 + \mu x_2 q_3 - \lambda x_1 q_1 - \mu x_3 q_2 = 0; \\ \lambda x_1 q_2 + \mu x_2 q_0 - \lambda x_2 q_1 - \mu x_0 q_2 = 0; \\ \mu x_0 q_3 + \lambda x_3 q_1 - \mu x_3 q_0 - \lambda x_1 q_3 = 0. \end{cases} \quad (10)$$

В этой системе первое уравнение совпадает со вторым, а третье — с четвертым. Легко установить, что все решения системы (10) описываются формулой

$$\mathbf{X} = (x_0, x_1, x_2, x_3) = \left(d, \frac{\mu q_2 d + (\lambda q_1 - \mu q_0) h}{\lambda q_2}, h, \frac{q_3}{q_2} h \right), \quad (11)$$

где d, $h = 0,1, \ldots, p-1$. Множество (11) содержит p^2 различных векторов, включая нулевой элемент (0, 0, 0, 0) и единицу $(\mu^{-1}, \lambda^{-1}, 0, 0)$ алгебры.

Утверждение 1. Любые два вектора **V** и **W** из множества (11) являются перестановочными, т. е. для них выполняется условие $\mathbf{W} \circ \mathbf{V} = \mathbf{V} \circ \mathbf{W}$.

Доказательство. Выполняется непосредственной проверкой с использованием формулы (11) для двух произвольных пар значений (d_1, h_1) и (d_2, h_2).

4. Свойства шестимерной НКАА

Характерной особенностью шестимерной НКАА, задаваемой табл. 2, является наличие в ней большого множества глобальных левосторонних единиц. Для нахождения формулы, описывающей это множество, следует рассмотреть векторное уравнение $\mathbf{X} \circ \mathbf{A} = \mathbf{A}$ при некотором фиксированном шестимерном векторе $\mathbf{A} = (a_0, a_1, a_2, a_3, a_4, a_5)$. С использованием табл. 2 указанное векторное уравнение сводится к следующей системе из шести линейных уравнений с неизвестными координатами вектора $\mathbf{X} = (x_0, x_1, x_2, x_3, x_4, x_5)$:

$$\begin{cases} (x_0 + \lambda x_4)a_0 + x_2 a_3 = a_0; \\ (\lambda x_1 + x_3)a_4 + x_5 a_1 = a_1; \\ (x_0 + \lambda x_4)a_2 + x_2 a_5 = a_2; \\ (\lambda x_1 + x_3)a_0 + x_5 a_3 = a_3; \\ (x_0 + \lambda x_4)a_4 + x_2 a_1 = a_4; \\ (\lambda x_1 + x_3)a_2 + x_5 a_5 = a_5. \end{cases} \quad (12)$$

Используя замену переменных по формулам $u_1 = x_0 + \lambda x_4$ и $u_2 = \lambda x_1 + x_3$, систему (12) можно представить в виде двух независимых систем из трех линейных уравнений:

$$\begin{cases} u_1 a_0 + x_2 a_3 = a_0; \\ u_1 a_2 + x_2 a_5 = a_2; \\ u_1 a_4 + x_2 a_1 = a_4; \end{cases} \quad (13)$$

$$\begin{cases} u_2 a_4 + x_5 a_1 = a_1; \\ u_2 a_0 + x_5 a_3 = a_3; \\ u_2 a_2 + x_5 a_5 = a_5. \end{cases} \quad (14)$$

Для произвольного вектора $\mathbf{A}=(a_0, a_1, a_2, a_3, a_4, a_5)$ система (13) имеет решение $(u_1, x_2) = (1, 0)$, а система (14) — решение $(u_2, x_5) = (0, 1)$. Путем обратной замены переменных получаем следующую формулу, описывающую множество p^2 глобальных левосторонних единиц $\mathbf{L} = (l_0, l_1, l_2, l_3, l_4, l_5)$, содержащихся в рассматриваемой НКАА:

$$\begin{aligned} &L = (d, h, 0, -\lambda h, \lambda^{-1}(1-d), 1), \\ &d, h = 0, 1, \ldots, p-1. \end{aligned} \quad (15)$$

Алгебра содержит локальные правосторонние единицы, которые могут быть вычислены из векторного уравнения $\mathbf{A} \circ \mathbf{X} = \mathbf{A}$, которое сводится к следующей тройке независимых систем из двух линейных уравнений с парами неизвестных значений (x_0, x_3), (x_1, x_4) и (x_2, x_5) соответственно:

$$\begin{cases} (a_0 + \lambda a_4)x_0 + a_2 x_3 = a_0; \\ (\lambda a_1 + a_3)x_0 + a_5 x_3 = a_3; \end{cases} \quad (16)$$

$$\begin{cases} a_5 x_1 + (\lambda a_1 + a_3)x_2 = a_1; \\ a_2 x_1 + (a_0 + \lambda a_4)x_2 = a_4; \end{cases} \quad (17)$$

$$\begin{cases} (a_0 + \lambda a_4)x_2 + a_2 x_5 = a_2; \\ (\lambda a_1 + a_3)x_2 + a_5 x_5 = a_5. \end{cases} \quad (18)$$

Главные определители систем (16), (17) и (18) равны $\Delta^{(16)} = -\Delta^{(17)} = \Delta^{(18)} = a_5(a_0 + \lambda a_4) - a_2(\lambda a_1 + a_3)$. Если координаты вектора **A** удовлетворяют условию $\Delta_A = \Delta^{(16)} \neq 0$, то к вектору **A** относится единственная локальная правосторонняя единица $\mathbf{R}_A = (r_0, r_1, r_2, r_3, r_4, r_5)$, координаты которой могут быть вычислены по следующим формулам:

$$r_0 = \frac{a_0 a_5 - a_2 a_3}{\Delta^{(16)}}; \; r_1 = \frac{a_0 a_1 - a_3 a_4}{\Delta^{(17)}};$$

$$r_1 = \frac{a_0 a_1 - a_3 a_4}{\Delta^{(17)}}; \; r_2 = 0;$$

$$r_3 = \frac{\lambda(a_3 a_4 - a_1 a_0)}{\Delta^{(16)}}; \; r_4 = \frac{a_4 a_5 - a_1 a_2}{\Delta^{(17)}};$$

$$r_5 = 1.$$

Легко доказать, что единица $\mathbf{R}_A$, относящаяся к произвольному вектору **A**, содержится в множестве глобальных левосторонних единиц (15). Следующие два утверждения достаточно очевидны.

Утверждение 2. Локальная правосторонняя единица $\mathbf{R}_A$ одновременно является локальной двухсторонней единицей $\mathbf{E}_A$, относящейся к вектору **A**.

Очевидно, что вектор **A** является обратимым относительно единицы $\mathbf{E}_A$, которая является единичным вектором конечной циклической группы, генерируемой степенями вектора **A**. Поэтому вектор **A** называется локально обратимым, если он удовлетворяет условию $\Delta_A = \Delta^{(16)} \neq 0$, т. е. условию

$$a_5(a_0 + \lambda a_4) - a_2(\lambda a_1 + a_3) \neq 0 \quad (19)$$

Утверждение 3. Пусть вектор **L** — глобальная левосторонняя единица. Тогда отображение алгебры, задаваемое формулой $\varphi_L(\mathbf{X}) = \mathbf{X} \circ \mathbf{L}$, где **X** пробегает все значения в рассматриваемой алгебре, является гомоморфизмом.

Доказательство. Для двух произвольных векторов $\mathbf{X}_1$ и $\mathbf{X}_2$ имеем:

$$\varphi_L(\mathbf{X}_1 \circ \mathbf{X}_2) = (\mathbf{X}_1 \circ \mathbf{X}_2) \circ \mathbf{L} =$$
$$= (\mathbf{X}_1 \circ \mathbf{L}) \circ (\mathbf{X}_2 \circ \mathbf{L}) = \varphi_L(\mathbf{X}_1) \circ \varphi_L(\mathbf{X}_2);$$
$$\varphi_L(\mathbf{X}_1 + \mathbf{X}_2) = (\mathbf{X}_1 + \mathbf{X}_2) \circ \mathbf{L} =$$
$$= (\mathbf{X}_1 \circ \mathbf{L}) + (\mathbf{X}_2 \circ \mathbf{L}) = \varphi_L(\mathbf{X}_1) + \varphi_L(\mathbf{X}_2).$$

Утверждение 4. Число локально обратимых векторов, содержащихся в рассматриваемой шестимерной алгебре, равно значению $\Omega = p^3(p-1)(p^2-1)$..

Доказательство. Найдем число η необратимых векторов. Координаты необратимого вектора $\mathbf{A} = (a_0, a_1, a_2, a_3, a_4, a_5)$ удовлетворяют условию $a_5(a_0 + \lambda a_4) - a_2(\lambda a_1 + a_3) = 0$. Число решений последнего уравнения с неизвестными a_0, a_1, a_2, a_3, a_4 и a_5 равно искомому значению η. В случае $a_5 \neq 0$ (имеем $p-1$ различных значений a_5) для каждого возможного набора значений координат a_0, a_1, a_3 и a_4 (p^4 вариантов) имеется единственное значение a_2, удовлетворяющее рассматриваемому уравнению, т. е. в случае $a_5 \neq 0$ имеется $p^4(p-1)$ необратимых векторов. В случае $a_5 = 0$ координаты a_0 и a_4 могут принимать произвольные значения (p^2 вариантов) при обязательном выполнении условия $a_2(\lambda a_1 + a_3) = 0$. Последнее дает $p^2 + p(p-1)$ вариантов, а в случае $a_5 = 0$ всего имеем $2p^4 - p^3$ необратимых векторов. Принимая во внимание оба рассмотренных случая, получаем $\eta = p^5 + p^4 - p^3$. В алгебре всего содержится p^6 различных векторов, из которых следует вычесть необратимые векторы, что дает следующее число локально обратимых векторов: $\Omega = p^6 - \eta = p^3(p-1)(p^2-1)$.

Пусть задана локальная правосторонняя единица $\mathbf{R}_A$. Она одновременно является и локальной двухсторонней единицей $\mathbf{E}_A$, относящейся к вектору $\mathbf{A}$. Очевидно, что полное подмножество векторов, относящихся к $\mathbf{E}_A$, составляет конечную группу Г с групповой единицей $\mathbf{E}_A$. Умножая каждый элемент группы Г на произвольную фиксированную глобальную левостороннюю единицу $\mathbf{L}$, получаем другую конечную группу, изоморфную с Г, единицей которой является вектор $\mathbf{R} \circ \mathbf{L} = \mathbf{L}$ (см. утверждение 3). Таким образом, каждая глобальная левосторонняя единица $\mathbf{L}$ задает существование в алгебре уникальной конечной группы с единицей $\mathbf{L}$. Поскольку имеется p^2 различных глобальных левосторонних единиц, то рассматриваемая шестимерная алгебра содержит p^2 различных изоморфных групп, порядок которых равен одному и тому же значению Ω'. Так как каждый локально обратимый вектор содержится только в одной из указанных групп, имеем $\Omega = p^2\Omega'$, откуда получаем формулу для порядка конечных групп, содержащихся в алгебре:

$$\Omega' = p(p-1)(p^2-1). \qquad (20)$$

Утверждение 5. Пусть $\mathbf{A} \circ \mathbf{B} = \mathbf{L}'$, где $\mathbf{L}'$ — глобальная левосторонняя единица. Тогда для произвольного натурального числа t выполняется равенство $\mathbf{A}^t \circ \mathbf{B}^t = \mathbf{L}'$.

Доказательство:

$$\mathbf{A}^t \circ \mathbf{B} = \mathbf{A}^{t-1} \circ (\mathbf{A} \circ \mathbf{B}) \circ \mathbf{B}^{t-1} = \mathbf{A}^{t-1} \circ \mathbf{B}^{t-1} =$$
$$= \mathbf{A}^{t-2} \circ (\mathbf{A} \circ \mathbf{B}) \circ \mathbf{B}^{t-2} = \mathbf{A}^{t-2} \circ \mathbf{B}^{t-2} = \mathbf{A} \circ \mathbf{B} = \mathbf{L}'.$$

Утверждение 6. Пусть $\mathbf{A} \circ \mathbf{B} = \mathbf{L}'$ и t — произвольнее натуральное число. Тогда формула $\psi_{L'} = \mathbf{B} \circ \mathbf{X} \circ \mathbf{A}$, где $\mathbf{X}$ пробегает все значения в рассматриваемой алгебре, является гомоморфизмом.

Доказательство. Для двух произвольных векторов $\mathbf{X}_1$ и $\mathbf{X}_2$ имеем:

$$\psi_{L'}(\mathbf{X}_1 \circ \mathbf{X}_2) = \mathbf{B} \circ (\mathbf{X}_1 \circ \mathbf{X}_2) \circ \mathbf{A} =$$
$$= \mathbf{B} \circ (\mathbf{X}_1 \circ \mathbf{L}' \circ \mathbf{X}_2) \circ \mathbf{A} =$$
$$= (\mathbf{B} \circ \mathbf{X}_1 \circ \mathbf{A}) \circ (\mathbf{B} \circ \mathbf{X}_2 \circ \mathbf{A}) =$$
$$= \psi_{L'}(\mathbf{X}_1) \circ \psi_{L'}(\mathbf{X}_2);$$
$$\psi_{L'}(\mathbf{X}_1 + \mathbf{X}_2) = \mathbf{B} \circ (\mathbf{X}_1 + \mathbf{X}_2) \circ \mathbf{A} =$$
$$= (\mathbf{B} \circ \mathbf{X}_1 \circ \mathbf{A}) + (\mathbf{B} \circ \mathbf{X}_2 \circ \mathbf{A}) =$$
$$= \psi_{L'}(\mathbf{X}_1) + \psi_{L'}(\mathbf{X}_2).$$

При использовании рассматриваемой шестимерной НКАА в качестве алгебраического носителя схемы открытого согласования ключей, основанной на СЗДЛ, маскирование базовой операции возведения в степень можно осуществить с помощью операции гомоморфного отображения $\psi_{L'}(\mathbf{X})$ с использованием в качестве общей пары векторов $\mathbf{A}$ и $\mathbf{B}$, удовлетворяющих условию $\mathbf{A} \circ \mathbf{B} = \mathbf{L}'$. В этом случае секретной является конкретная модификация данной операции, определяемая выбором секретного значения степени t, задающей конкретное гомоморфное преобразование, определяемое формулой $\psi_{L'}(\mathbf{X}) = \mathbf{B}^t \circ \mathbf{X} \circ \mathbf{A}^t$.

5. Схема открытого согласования ключа на основе СЗДЛ в четырехмерной алгебре

При использовании четырехмерной НКАА с операцией умножения, задаваемой по табл. 1, в качестве общих параметров схемы открытого согласования ключей выбирается 1) характеристика простого конечного поля $GF(p)$,

равная значению $p = 2q - 1$, где q — 256-битовое простое число; 2) обратимый вектор $\mathbf{N} = (n_0, n_1, n_2, n_3)$, порядок которого равен q; 3) обратимый вектор $\mathbf{Q}$, удовлетворяющий условию $\mathbf{N} \circ \mathbf{Q} \neq \mathbf{Q} \circ \mathbf{N}$, координаты которого q_0, q_1, q_2 и q_3 используются для выполнения вычисления случайных векторов $\mathbf{X}$ по формуле (11).

Генерация открытого ключа пользователя выполняется следующим образом:

1. Пользователь выбирает случайные натуральные числа $x < q$, $d < p$ и $h < p$.

2. По формуле (11) и случайно выбранным значениям d и h вычисляет вектор $\mathbf{X}$.

3. Вычисляет открытый ключ $\mathbf{Y} = \mathbf{X} \circ \mathbf{N}^x \circ \mathbf{X}^{-1}$.

Личным секретным ключом пользователя является число x и вектор $\mathbf{X}$. Общий секретный ключ двух пользователей формируется следующим образом. Первый пользователь, используя свой секретный ключ (x_1, $\mathbf{X}_1$) и открытый ключ $\mathbf{Y}_2$ второго пользователя, вычисляет вектор

$$\begin{aligned}\mathbf{Z}_1 &= \mathbf{X}_1 \circ \mathbf{Y}_2^{x_1} \circ \mathbf{X}_1^{-1} = \\ &= \mathbf{X}_1 \circ (\mathbf{X}_2 \circ \mathbf{N}^{x_2} \circ \mathbf{X}_2^{-1})^{x_1} \circ \mathbf{X}_1^{-1} = \\ &= \mathbf{X}_1 \circ \mathbf{X}_2 \circ \mathbf{N}^{x_2 x_1} \circ \mathbf{X}_2^{-1} \circ \mathbf{X}_1^{-1}.\end{aligned}$$

Второй пользователь, используя свой секретный ключ (x_2, $\mathbf{X}_2$) и открытый ключ $\mathbf{Y}_1$ первого пользователя, вычисляет вектор

$$\begin{aligned}\mathbf{Z}_2 &= \mathbf{X}_2 \circ \mathbf{Y}_1^{x_2} \circ \mathbf{X}_2^{-1} = \\ &= \mathbf{X}_2 \circ (\mathbf{X}_1 \circ \mathbf{N}^{x_1} \circ \mathbf{X}_1^{-1})^{x_2} \circ \mathbf{X}_2^{-1} = \\ &= \mathbf{X}_2 \circ \mathbf{X}_1 \circ \mathbf{N}^{x_1 x_2} \circ \mathbf{X}_1^{-1} \circ \mathbf{X}_2^{-1}.\end{aligned}$$

Учитывая перестановочность векторов $\mathbf{X}_1$ и $\mathbf{X}_2$, легко показать, что выполняется условие $\mathbf{Z}_1 = \mathbf{Z}_2$, т. е. оба пользователя вычисляют один и тот же вектор, который служит общим секретным ключом, согласованным по открытому каналу связи. При практическом использовании этой криптосхемы предполагается применение механизмов проверки подлинности открытых ключей пользователей, что является стандартным условием применения протоколов данного типа. Например, пользователи пересылают друг другу цифровые сертификаты, содержащие значения их открытых ключей и подписанные цифровой подписью удостоверяющего центра.

6. Схема открытого согласования ключа на основе СЗДЛ в шестимерной алгебре

При использовании шестимерной НКАА с операцией умножения, задаваемой по табл. 2, в качестве общих параметров схемы открытого согласования ключей выбирается 1) характеристика простого конечного поля $GF(p)$, равная значению $p = 2q - 1$, где q — 256-битовое простое число; 2) локально обратимый вектор $\mathbf{N} = (n_0, n_1, n_2, n_3)$, порядок которого равен q; 3) пара локально обратимых векторов $\mathbf{A}$ и $\mathbf{B}$, удовлетворяющих условию $\mathbf{A} \circ \mathbf{B} = \mathbf{L}$, , где $\mathbf{L}$ — глобальная левосторонняя единица.

Генерация открытого ключа пользователя выполняется следующим образом:

1. Пользователь выбирает случайные натуральные числа $x < q$ и $t < q$.

2. Вычисляет открытый ключ $\mathbf{Y} = \mathbf{B}^t \circ \mathbf{N}^x \circ \mathbf{A}^t$.

Личным секретным ключом пользователя является пара чисел x и t. Общий секретный ключ двух пользователей формируется следующим образом. Первый пользователь, используя свой секретный ключ (x_1, t_1) и открытый ключ $\mathbf{Y}_2$ второго пользователя, вычисляет вектор

$$\begin{aligned}\mathbf{Z}_1 &= \mathbf{B}^{t_1} \circ \mathbf{Y}_2^{x_1} \circ \mathbf{A}^{t_1} = \\ &= \mathbf{B}^{t_1} \circ (\mathbf{B}^{t_2} \circ \mathbf{N}^{x_2} \mathbf{A}^{t_2})^{x_1} \circ \mathbf{A}^{t_1} = \\ &= \mathbf{B}^{t_1 + t_2} \circ \mathbf{N}^{x_2 x_1} \circ \mathbf{A}^{t_2 + t_1}\end{aligned}$$

Второй пользователь, используя свой секретный ключ (x_2, t_2) и открытый ключ $\mathbf{Y}_1$ первого пользователя, вычисляет вектор

$$\begin{aligned}\mathbf{Z}_2 &= \mathbf{B}^{t_2} \circ \mathbf{Y}_1^{x_2} \circ \mathbf{A}^{t_2} = \\ &= \mathbf{B}^{t_2} \circ (\mathbf{B}^{t_1} \circ \mathbf{N}^{x_1} \mathbf{A}^{t_1})^{x_2} \circ \mathbf{A}^{t_2} = \\ &= \mathbf{B}^{t_2 + t_1} \circ \mathbf{N}^{x_1 x_2} \circ \mathbf{A}^{t_1 + t_2} = \mathbf{Z}_1.\end{aligned}$$

Таким образом, оба пользователя вычисляют один и тот же вектор, который служит общим секретным ключом, согласованным по открытому каналу связи.

Заключение

Предложены новые реализации схем открытого согласования ключа, основанные на вычислительной трудности скрытой задачи дискретного логарифмирования, в которых для повышения их производительности в качестве алгебраических носителей используются НКАА с операцией умножения, заданной по прореженным ТУБВ. Первая предложенная криптосхема использует вычисления в НКАА с глобальной двухсторонней единицей и новый механизм формирования маскирующей операции. Вторая предложенная криптосхема реализует маскирующие операции, ранее предложенные в работе [14], и отличается применением шестимерных НКАА, заданных по прореженной ТУБВ. Ранее прореженные ТУБВ использовались для задания четырехмерных НКАА.

Построение шестимерных НКАА, заданных по прореженной ТУББ, выполнено впервые.

Дальнейшее развитие протоколов открытого согласования ключа, основанных на СЗДЛ, представляет интерес в направлении применения новых типов маскирующих операций и использования в качестве алгебраических носителей НКАА, заданных над конечными расширениями двоичного поля $GF(2^s)$.

Список литературы

1. **Proceedings** of the 7th International Workshop on Post-Quantum Cryptography, PQCrypto 2016. Fukuoka, Japan, February 24-26, 2016 // Lecture Notes in Computer Science (LNCS) series. Springer, 2016. Vol. 9606. 270 p.
2. **Federal Register**. Announcing Request for Nominations for Public-Key Post-Quantum Cryptographic Algorithms. URL: https://www.gpo.gov/fdsys/pkg/FR-2016-12-20/pdf/2016-30615.pdf (дата обращения 16.04.2020).
3. **Post-Quantum** Cryptography. 9th International Conference, PQCrypto 2018, Fort Lauderdale, FL, USA, April 9-11, 2018, Proceedings // Lecture Notes in Computer Science series. Springer, 2018. Vol. 10786.
7. **Post-Quantum** Cryptography. Proceedings of the 10th International Conference, PQCrypto 2019, Chongqing, China, May 8—10, 2019 // Lecture Notes in Computer Science. 2019. Vol. 11505. 420 p.
5. **Shor P. W.** Polynomial-time algorithms for prime factorization and discrete logarithms on quantum computer // SIAM Journal of Computing. 1997. Vol. 26. P. 1484—1509.
6. **Smolin J. A., Smith G., Vargo A.** Oversimplifying quantum factoring // Nature. 2013. Vol. 499, N. 7457. P. 163—165.
7. **Yan S. Y.** Quantum Computational Number Theory. Springer, 2015. 252 p.
8. **Yan S. Y.** Quantum Attacks on Public-Key Cryptosystems. Springer, 2014. 207 p.
9. **Ekert A., Jozsa R.** Quantum computation and Shor's factoring algorithm // Rev. Mod. Phys. 1996. Vol. 68. P. 733.
10. **Jozsa R.** Quantum algorithms and the fourier transform // Proc. Roy. Soc. London Ser A. 1998. Vol. 454. P. 323-337.
11. **Moldovyan D. N.** Non-Commutative Finite Groups as Primitive of Public-Key Cryptoschemes // Quasigroups and Related Systems. 2010. Vol. 18. P. 165—176.
12. **Кузьмин А. С., Марков В. Т., Михалев А. А., Михалев А. В., Нечаев А. А.** Криптографические алгоритмы на группах и алгебрах // Фундаментальная и прикладная математика. 2015. Т. 20, № 1. С. 205—222.
13. **Kuzmin A. S., Markov V. T., Mikhalev A. A., Mikhalev A. V., Nechaev A. A.** Cryptographic Algorithms on Groups and Algebras // Journal of Mathematical Sciences. 2017. Vol. 223, N. 5. P. 629—641.
14. **Moldovyan D. N.** Post-quantum public key-agreement scheme based on a new form of the hidden logarithm problem // Computer Science Journal of Moldova. 2019. V.27, N. 1(79). P. 56—72.
15. **Moldovyan A. A., Moldovyan N. A.** Post-quantum signature algorithms based on the hidden discrete logarithm problem // Computer Science Journal of Moldova. 2018. V. 26, N. 3(78). P. 301—313.
16. **Молдовян А. А., Молдовян Н. А.** Новые формы задания скрытой задачи дискретного логарифмирования // Труды СПИИРАН. 2019. № 2 (18). С. 504—529.
17. **Moldovyan A. A., Moldovyan D. N., Moldovyan N. A.** Post-quantum commutative encryption algorithm // Computer Science Journal of Moldova. 2019. V. 27, N. 3(81). P. 299—317.
18. **Абросимов И. К., Ковалева И. В., Молдовян Н. А.** Постквантовый протокол бесключевого шифрования // Вопросы защиты информации. 2017. № 3. С. 3—13.
19. **Moldovyan N. A., Moldovyan A. A.** Finite Non-commutative Associative Algebras as carriers of Hidden Discrete Logarithm Problem // Bulletin of the South Ural State University. Ser. Mathematical Modelling, Programming & Computer Software (Bulletin SUSU MMCS). 2019. Vol. 12, N. 1. P. 66—81.
20. **Moldovyan A. A.** General Method for Defining Finite Non-commutative Associative Algebras of Dimension m>1 // Buletinul Academiei de Stiinte a Republicii Moldova. Matematica. 2018. N. 2 (87). P. 95—100.
21. **Moldovyan N. A.** Unified Method for Defining Finite Associative Algebras of Arbitrary Even Dimensions // Quasigroups and Related Systems. 2018. Vol. 26, N. 2. P. 263—270.
22. **Moldovyan D. N.** A unified method for setting finite non-commutative associative algebras and their properties // Quasigroups and Related Systems. 2019. Vol. 27, N. 2. P. 293—308.
23. **Молдовян Н. А., Абросимов И. К.** Схема постквантовой электронной цифровой подписи на основе усиленной формы скрытой задачи дискретного логарифмирования // Вестник Санкт-Петербургского университета. Прикладная математика. Информатика. Процессы управления. 2019. Т. 15, Вып. 2. С. 212—220.
24. **Молдовян Н. А., Абросимов И. К.** Постквантовые протоколы цифровой подписи на основе скрытой задачи дискретного логарифмирования // Вопросы защиты информации. 2019. № 2. С. 23—32.
25. **Молдовян А. А., Молдовян Д. Н.** Постквантовая схема ЭЦП на основе скрытой задачи дискретного логарифмирования в четырехмерной конечной алгебре // Вопросы защиты информации. 2019. № 2. С. 18—22.

R. S. Fahrutdinov, Head of Laboratory, e-mail: fahr@cobra.ru,
A. Yu. Mirin, Senior Researcher, e-mail: mirin@cobra.ru,
D. N. Moldovyan, Researcher, e-mail: mdn.spectr@mail.ru,
A. A. Kostina, Researcher, e-mail: anna-kostina1805@mail.ru,
St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences,
St. Petersburg, 199178, Russian Federation

Public Key-Agreement Schemes Based on the Hidden Discrete Logarithm Problem

There is considered the problem of increasing the performance of the public kley-agreement schemes based on the computational complexity of the hidden discrete logarithm problem defined in finite non-commutative associative algebras of various types. To increase the rate of cryptoschemes of the said type, it is proposed to use algebras as their algebraic support, in which the associative multiplication operation is specified using sparse multiplication tables of basis vectors. In framework of this method the rate increase is achieved by a significant reduction in the number of multiplications in the finite field, over which the algebra is specified, which are necessary to perform one multiplication operation in the algebra. The principal realizability of this method has

been demonstrated for cases of four-dimensional and six-dimensional algebras, for which the sparse tables are given that specify the associative multiplication operation and providing two-times reduction of the number of multiplications in the field. Another proposed way to increase the rate is to specify the procedure for generating permutable key elements in the form of a computational procedure performed according to specially derived mathematical formulas, free from the operation of exponentiation to a large-size integer power. The second method is based on the idea of defining a set of mutually permutable vectors of a finite non-commutative associative algebra, described by a fairly compact mathematical formula. Moreover, the latter defines a procedure for calculating a random vector from the indicated set of vectors, which has significantly lower computational complexity compared to the exponentiation operation used in well-known cryptoschemes of the considered type to generate random pairs of permutable vectors. The potential feasibility of the second method is demonstrated by the derivation of the indicated formula for a four-dimensional algebra given by sparse multiplication tables of basis vectors. Specific public key-agreement cryptoschemes have been developed that implement the developed methods for increasing performance, which are of interest for practical use as post-quantum public key-agreement schemes. To further increase the performance of cryptoschemes of the considered type, it is proposed to use the algebras set over finite extensions of a binary field.

Keywords: *information protection, cryptography, public key-agreement, discrete logarithm problem, finite associative algebra, non-commutative algebra, global unit, local unit, left-sided unit*

Acknowledgements: This work was supported by the budget theme № 0060-2019-0010.

DOI: 10.17587/it.26.577-585

References

1. **Proceedings** of the 7th International Workshop on Post-Quantum Cryptography, PQCrypto 2016. Fukuoka, Japan, February 24—26, 2016, *Lecture Notes in Computer Science (LNCS) series*, Springer, 2016, vol. 9606, 270 p.

2. **Federal** Register. Announcing Request for Nominations for Public-Key Post-Quantum Cryptographic Algorithms, available at: https://www.gpo.gov/fdsys/pkg/FR-2016-12-20/pdf/2016-30615.pdf (date of access 16.04.2020).

3. **Post-Quantum** Cryptography. 9th International Conference, PQCrypto 2018, Fort Lauderdale, FL, USA, April 9—11, 2018, Proceedings, *Lecture Notes in Computer Science series*, Springer, 2018, vol. 10786.

4. **Post-Quantum** Cryptography. Proceedings of the 10th International Conference, PQCrypto 2019, Chongqing, China, May 8—10, 2019, *Lecture Notes in Computer Science*, 2019, vol. 11505, 420 p.

5. **Shor P. W.** Polynomial-time algorithms for prime factorization and discrete logarithms on quantum computer, *SIAM Journal of Computing*, 1997, vol. 26, pp. 1484—1509.

6. **Smolin J. A., Smith G., Vargo A.** Oversimplifying quantum factoring, *Nature*, 2013, vol. 499, no. 7457, pp. 163—165.

7. **Yan S. Y.** Quantum Computational Number Theory, Springer, 2015, 252 p.

8. **Yan S. Y.** Quantum Attacks on Public-Key Cryptosystems, Springer, 2014, 207 p.

9. **Ekert A., Jozsa R.** Quantum computation and Shor's factoring algorithm, *Rev. Mod. Phys.*, 1996, vol. 68, pp. 733.

10. **Jozsa R.** Quantum algorithms and the fourier transform, *Proc. Roy. Soc. London Ser A*, 1998, vol. 454, pp. 323—337.

11. **Moldovyan D. N.** Non-Commutative Finite Groups as Primitive of Public-Key Cryptoschemes, *Quasigroups and Related Systems,* 2010, vol. 18, pp. 165—176.

12. **Kuzmin A. S., Markov V. T., Mikhalev A. A., Mikhalev A. V., Nechaev A. A.** Cryptographic Algorithms on Groups and Algebras, *Fundamental'naja i prikladnaja matematika*, 2015, vol. 20, no. 1, pp. 205—222 (in Russian).

13. **Kuzmin A. S., Markov V. T., Mikhalev A. A., Mikhalev A. V., Nechaev A. A.** Cryptographic Algorithms on Groups and Algebras, *Journal of Mathematical Sciences,* 2017, vol. 223, no. 5, pp. 629—641.

14. **Moldovyan D. N.** Post-quantum public key-agreement scheme based on a new form of the hidden logarithm problem, *Computer Science Journal of Moldova,* 2019, vol.27, no. 1(79), pp. 56—72.

15. **Moldovyan A. A., Moldovyan N. A.** Post-quantum signature algorithms based on the hidden discrete logarithm problem, *Computer Science Journal of Moldova,* 2018, vol.26, no. 3(78), pp. 301—313.

16. **Moldovyan A. A., Moldovyan N. A.** New Forms of Defining the Hidden Discrete Logarithm Problem, *Trudy SPIIRAN,* 2019, no. 2 (18), pp. 504—529 (in Russian).

17. **Moldovyan A. A., Moldovyan D. N., Moldovyan N. A.** Post-quantum commutative encryption algorithm, *Computer Science Journal of Moldova,* 2019, vol.27, no. 3(81), pp. 299—317.

18. **Abrosimov I. K., Kovaleva I. V., Moldovyan N. A.** Post-quantum protocol of keyless encryption, *Voprosy Zashhity Informacii,* 2017, no. 3, pp. 3—13 (in Russian).

19. **Moldovyan N. A., Moldovyan A. A.** Finite Non-commutative Associative Algebras as carriers of Hidden Discrete Logarithm Problem, *Bulletin of the South Ural State University. Ser. Mathematical Modelling, Programming & Computer Software (Bulletin SUSU MMCS)*, 2019, vol. 12, no. 1, pp. 66—81.

20. **Moldovyan A. A.** General Method for Defining Finite Non-commutative Associative Algebras of Dimension m>1, *Buletinul Academiei de Stiinte a Republicii Moldova. Matematica,* 2018, no. 2 (87), pp. 95—100.

21. **Moldovyan N. A.** Unified Method for Defining Finite Associative Algebras of Arbitrary Even Dimensions, *Quasigroups and Related Systems,* 2018, vol. 26, no. 2, pp. 263—270.

22. **Moldovyan D. N.** A unified method for setting finite non-commutative associative algebras and their properties, *Quasigroups and Related Systems,* 2019, vol. 27, no. 2, pp. 293—308.

23. **Moldovyan N. A., Abrosimov I. K.** Post-quantum digital signature schemes based on the enhanced form of the hidden discrete logarithm problem, *Vestnik Sankt-Peterburgskogo universiteta. Prikladnaja Matematika. Informatika. Processy Upravlenija,* 2019, vol. 15, iss. 2, pp. 212—220 (in Russian).

24. **Moldovyan N. A., Abrosimov I. K.** Post-quantum digital signature protocols based on the hidden discrete logarithm problem, *Voprosy Zashhity Informacii,* 2019, no. 2, pp. 23—32 (in Russian).

25. **Moldovyan A. A., Moldovyan D. N.** Post-quantum digital signature schemes based on the hidden discrete logarithm problem in four-dimensional finite algebra, *Voprosy Zashhity Informacii,* 2019, no. 2, pp. 18—22 (in Russian).

УДК 004.932 DOI: 10.17587/it.26.586-593

А. В. Савченко, д-р техн. наук, проф., e-mail: avsavchenko@hse.ru,
И. С. Гречихин, аспирант, ст. преподаватель, e-mail: igrechikhin@hse.ru,
Национальный исследовательский университет Высшая школа экономики, Нижний Новгород

Детектирование специализированных категорий объектов на фотографиях в мобильных устройствах на основе многозадачной нейросетевой модели[1]

Предложен метод детектирования категорий нескольких различных видов объектов на фотографиях в мобильных устройствах. Вначале с использованием известных нейросетевых детекторов выделяются искомые объекты. Их характерные признаки извлекаются с помощью многозадачной нейросетевой модели с несколькими выходными слоями — по одному на каждый вид объекта. Представлены экспериментальные результаты для распознавания пород собак и кошек и группировки фотографий одного и того же животного.

Ключевые слова: *обработка изображений, сверточные нейронные сети, мобильные системы, распознавание пород животных, иерархическая кластеризация*

Введение

Задача определения предпочтений пользователей мобильных устройств в настоящее время становится все более актуальной в связи с непрерывным развитием рекомендательных систем. Один из вариантов ее решения связан с обработкой фотографий, сделанных самим пользователем. Как отмечено в работе [1], наибольший интерес представляют алгоритмы детектирования объектов на изображениях (предметов интерьера, видов еды, транспорта, спортивных принадлежностей, музыкальных инструментов и т.п.). Нередко требуется получить более специализированную информацию о предпочтении, в частности, определить подкатегории некоторых значимых классов (марки автомобилей, породы животных). К сожалению, современные наборы данных, предназначенные для обучения нейросетевых детекторов, не содержат данных о подкатегориях, либо существующих в этих наборах подкатегорий недостаточно для надежного обнаружения соответствующих им объектов.

Стоит отметить, что для многих важных подкатегорий доступны специализированные наборы данных, которые можно использовать для обучения классификаторов, например, глубоких сверточных нейронных сетей (СНС) [2]. Поэтому в настоящей работе используется двухэтапная процедура обнаружения объектов, в которой вначале для обнаружения более общих видов объектов применяются традиционные нейросетевые детекторы [3], а далее для нахождения специализированных подкатегорий применяется СНС. При этом особенность предлагаемого метода состоит в применении единой многозадачной СНС с несколькими выходами (по одному — для каждого вида объекта) [4]. В результате применения такого подхода можно не только снизить вычислительную сложность за счет отказа от применения нескольких СНС, но и использовать выходы промежуточных слоев сети в качестве характерных признаков анализируемых объектов, например, для группировки различных фото-

[1] Статья подготовлена в ходе проведения исследования (№ 19-04-004) в рамках Программы "Научный фонд Национального исследовательского университета "Высшая школа экономики" (НИУ ВШЭ)" в 2019—2020 гг. и в рамках государственной поддержки ведущих университетов Российской Федерации "5-100".

графий одного и того же объекта. В качестве примера реализации такого подхода в работе рассматривается классификация и последующая кластеризация различных видов домашних животных (пород кошек и собак) [5, 6]. Полученные результаты и сделанные по ним выводы рассчитаны на широкий круг специалистов в области распознавания образов.

1. Постановка задачи

Задача анализа предпочтений по фотографиям состоит в том, чтобы по поступившему на вход фотоальбому выделить наиболее интересные для пользователя категории из заранее заданного списка [1]. Результатом анализа предпочтений можно считать частоты встречаемости объектов каждой категории в фотоальбоме. Если для каждой категории задано множество изображений, соответствующих данной категории объектов, а также данные об их местонахождении на изображении (обрамляющие прямоугольники или маска границ), можно решить задачу с помощью обучения одного из современных высокоточных нейросетевых детекторов [7]. Архитектуры Faster R-CNN [3] также используют СНС для создания карты признаков, но с их помощью определяются несколько (100...200) регионов, в которых могут содержаться потенциально интересные объекты. После этого на основании карты признаков и выделенных регионов предсказывается класс объекта. В совокупности такая архитектура обнаруживает объекты значительно точнее за счет снижения вычислительной эффективности. Детектор SSD (Single Shot Detector) использует карту признаков на выходе СНС для предсказания классов и положения объектов за один проход, а его модификация SSDLite [8] включает разделяемые по глубине (depth-separable) сверточные слои для снижения вычислительной сложности и затрат памяти, что делает их удобными для использования в мобильных устройствах. Среди моделей, осуществляющих детектирование за один проход, следует выделить RetinaNet [9], которая позволяет за счет специальной функции потерь (focal loss) достичь достаточно высоких показателей точности и вычислительной эффективности.

Прорывом в области создания СНС, приемлемых как для классификации, так и для извлечения карт признаков, подходящих для использования в детекторах SSD и Faster R-CNN, стали архитектуры ResNet и Inception [10], которые сумели достичь высокого качества классификации изображений на значительном по размеру наборе изображений и категорий объектов. Для использования в мобильных устройствах [1], где есть ограничения по объему памяти и процессорной мощности, необходимы более вычислительно эффективные архитектуры, такие как MobileNet [8].

2. Многозадачные нейронные сети для классификации подкатегорий

К сожалению, во многих случаях сбор необходимого для обучения детектора набора данных оказывается слишком сложным. В частности, основная трудность состоит в получении разметки, необходимой для обучения детектора. Для этого на каждом изображении из обучающей выборки требуется указать область искомого объекта, чаще всего, с помощью выделения обрамляющего прямоугольника. При этом для получения высокой точности требуются сотни размеченных примеров каждого класса, и чем больше различных категорий, тем больше должно быть примеров объектов каждой категории.

В таком случае можно воспользоваться двухэтапной процедурой, в которой вначале с помощью нейросетевого *детектора* находятся $N > 1$ более общих видов объектов (автомобиль, еда, музыкальный инструмент, домашнее животное), а потом для каждого n-го вида ($n = 1, 2, ..., N$) выделяются специализированные категории. Пусть для n-го вида имеются $C_n > 1$ категорий. По результатам детектирования находятся обрамляющие прямоугольники для объекта n-го вида, после чего для каждого такого прямоугольника из изображения вырезается часть, принадлежащая найденному объекту. Далее выделенный объект распознается с помощью отдельного (для каждого n) *классификатора*, например, СНС.

Конечно, в этом случае время принятия решений может увеличиться за счет появления второй СНС. Однако такие классификаторы можно обучить, используя набор фотографий каждой подкатегории объектов n-го вида, в котором не требуется указывать обрамляющие прямоугольники, что существенно упрощает процедуру сбора и разметки данных. Более того, постоянно развиваются методы дообучения СНС на сверхмалых обучающих выборках (даже с одним примером каждой категории) [11], в то время как обучение части сети-детектора, следующей после извлечения карты

признаков, все еще требует больших объемов обучающих данных.

В связи с тем, что в процессе принятия решений исходные фотографии подаются на вход нейросетевого детектора, для обучения классификаторов также должен использоваться не исходный набор, а части изображений, полученные с помощью аналогичной процедуры выделения обрамляющих прямоугольников на выходе обученного детектора. Рассмотрим подробнее различные способы построения классификаторов на основе СНС.

Наиболее простой вариант — обучить отдельные классификаторы для каждого вида объекта. В настоящий момент в рамках технологии переноса знаний (transfer learning) [2] наиболее часто для настройки классификатора применяется не доступное обучающее множество, а сверхбольшая коллекция дополнительно собранных изображений, например, ImageNet. Такая коллекция используется для обучения глубокой СНС, состоящей из нескольких чередующихся слоев свертки и подвыборки, выход которых поступает на вход последовательно соединенных полносвязных слоев. Выход последнего сверточного слоя является четырехмерным тензором, поэтому далее обычно добавляется слой глобального усреднения (global average pooling) по ширине и высоте, после чего его выход из $D \gg 1$ значений поступает на вход последнего полносвязного слоя, в котором и принимается решение в пользу одной из подкатегорий. Такую архитектуру можно рассматривать как применение логистической регрессии (последний слой СНС) для классификации вектора $\mathbf{x}$ из D характерных признаков, выделенных на предыдущих слоях. Поэтому обычно последний полносвязный слой заменяется на новый слой с C_n выходами $\mathbf{z}_c$ (по одному на каждую подкатегорию n-го вида), в котором с помощью слоя softmax оцениваются апостериорные вероятности p_c принадлежности входного объекта c-й подкатегории ($c = 1, 2, ..., C_n$). После этого происходит дообучение (fine-tuning) полученного таким образом нейросетевого классификатора для доступного n-го обучающегося множества [2].

Такой способ является наиболее приемлемым, если для каждого вида доступно репрезентативное обучающее множество, при этом сами типы объектов существенно отличаются друг от друга. К сожалению, затраты памяти линейно зависят от числа видов N, при этом точность обученного классификатора может оказаться достаточно низкой, если имеется обучающая выборка малого размера. При этом, если необходимо добавить новый $(N + 1)$-й вид объектов, придется заново обучать новый классификатор.

Для преодоления указанных недостатков может применяться единая СНС, обученная для одновременного решения нескольких задач. Например, можно объединить все подкатегории в одно обучающее множество, состоящее из $(C_1 + C_2 + ... + C_N)$ классов. При этом необходимо выполнить дополнительную постобработку результатов классификации: использовать оценки апостериорных вероятностей, соответствующих только виду объекта, найденного детектором. Такой подход позволяет поддерживать всего один классификатор (рис. 1), однако при его обучении изображения одного вида оказывают влияние на выходы, ответственные за подкатегории других видов, что может привести к снижению итоговой точности.

Поэтому наиболее приемлемым способом реализации многозадачной СНС является использование одной сети для извлечения вектора признаков $\mathbf{x}$, который подается на N выходных слоев (heads) — по одному для каждого вида объекта. Такой подход (рис. 2) позволяет создать разные классификаторы на базе одной общей архитектуры СНС, которая получает карты признаков их входных изображений и передает одному из выходов для классификации. К недостаткам такого подхода можно отнести усложнение процесса обучения: веса нейронной сети, передающие информацию от карт признаков к каждому из выходов, для более сбалансированного обучения модифицируются отдельно в рамках итеративной процедуры для подвыборок объединенного обучающего множества (mini-batch), каждая из которых включает только один вид объектов [4].

Отметим, что на практике описанные выше подходы могут комбинироваться в гибридные архитектуры, если есть несколько схожих ви-

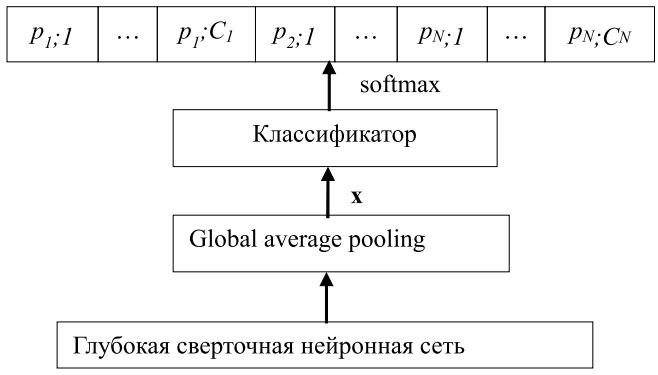


Рис. 1. Сверточная нейронная сеть с объединением всех подкатегорий

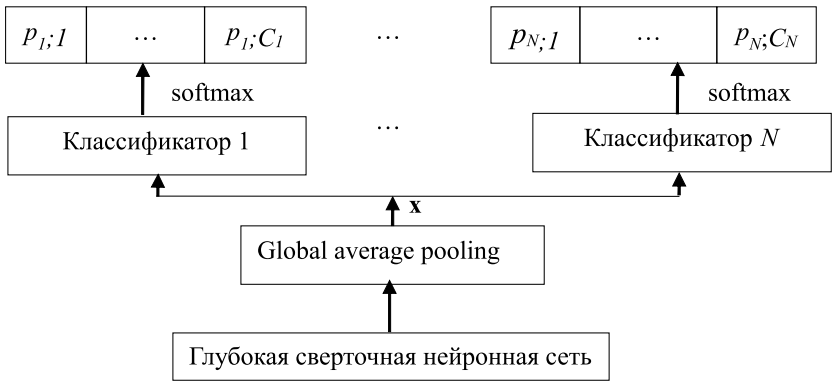


Рис. 2. Многозадачная сверточная нейронная сеть с N выходными слоями

дов объектов (например, породы котов и собак), которые существенно отличаются от других видов (например, автомобилей или видов еды). В таком случае использование единой базовой СНС для извлечения характерных признаков может оказаться неприемлемым, поэтому нужно использовать несколько независимых архитектур вида (рис. 2).

3. Предложенный подход

На рис. 3 представлена функциональная схема предлагаемой информационной системы извлечения предпочтений на основе детектирования категорий нескольких различных видов объектов. Здесь для каждой фотографии на первом этапе осуществляется детектирование общих видов объектов, специализированные категории которых на втором этапе предсказываются с помощью многозадачной СНС (см. рис. 2).

Кроме того, такая СНС применяется также для группировки фотографий идентичных объектов. Действительно, в галерее пользователя могут встречаться несколько фотографий одного и того же животного. Можно сделать предположение, что это животное представляет интерес для пользователя, например, может являться его домашним животным.

Для группировки объектов обычно применяются методы кластеризации [12]. Для того чтобы использовать алгоритмы кластеризации, каждый объект должен быть представлен некоторым числовым вектором. В частности, могут быть использованы вектор признаков **x** или вектор **p** апостериорных вероятностей классов p_c на выходе СНС. Так как число кластеров заранее не известно, в настоящей работе использовались методы агломеративной иерархической кластеризации и плотностной кластеризации (DBSCAN, HDBSCAN) [13], не требующие наличия информации о числе групп K. Первая группа методов создает иерархическую структуру кластеров, что может оказаться полезным при определении принадлежности кластеров. Для ее применения необходимо подобрать несколько гиперпараметров: мера близости, межкластерное расстояние и порог определения кластеров. Наилучшие параметры DBSCAN зависят от выборки и могут меняться в зависимости от набора векторов, поскольку этот алгоритм определяет кластеры одинаковой плотности. Метод HDBSCAN преодолевает это ограничение, однако он более вычислительно требователен [13].

Описанный алгоритм был реализован авторами в Android-приложении [14] анализа пользовательских предпочтений для мобильных устройств [1]. В приложении показываются фотографии из галереи пользователя, детектируются несколько видов объектов (автомобили, кошки и собаки), классифицируются марки автомобилей и породы домашних животных (рис. 4, см. третью сторону обложки). На странице со статистикой найденных категорий отмечаются как породы животных, так и статистика по найденным кластерам.

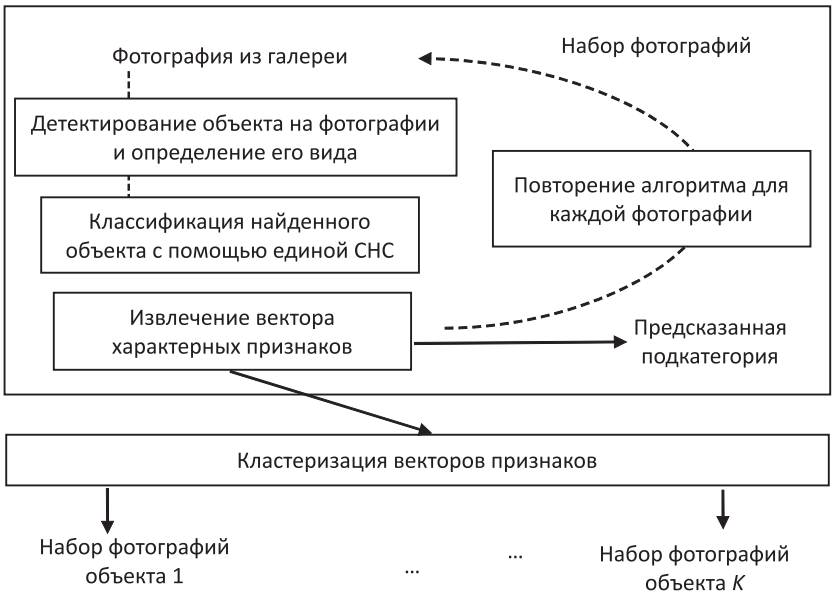


Рис. 3. Схема информационной системы извлечения предпочтений на основе детектирования категорий нескольких различных видов объектов

4. Вычислительный эксперимент

Эксперимент 1: классификации пород домашних животных. Для обучения моделей используются два набора данных (рис. 5, см. третью сторону обложки): Stanford Dogs Dataset [5] и Oxford-IIIT-Pet [6]. Набор Stanford Dogs Dataset содержит изображения $C_1 = 120$ пород собак (150...200 изображений для каждой породы). Oxford-IIIT-Pet содержит породы кошек и собак, для работы были взяты только изображения $C_2 = 12$ пород кошек (примерно 200 изображений на каждую породу).

При подготовке к обучению 70 наугад выбранных изображений каждой категории были помещены в обучающую выборку, а оставшиеся изображения были использованы для тестирования. Применялись различные архитектуры глубоких СНС — как высокоточные вычислительно сложные модели (Inception V3, Inception ResNet v2, ResNet-50, ResNet-101), так и легковесные нейронные сети MobileNet v1, MobileNet v2. Для практического исследования из библиотеки Keras 2.2.5 были взяты СНС, предварительно обученные для классификации изображений из базы данных ImageNet-1000.

В эксперименте сопоставляли два описанных выше способа организации многозадачной СНС: объединение всех подкатегорий (см. рис. 1) для разных видов животных (кошек и собак с общим числом 132 класса) и наличие двух выходов (см. рис. 2) — по одному для каждого вида животных. Каждый классификатор обучался в течение 120 эпох с помощью оптимизатора Adam, при этом в течение первых 20 эпох обучались только веса, связанные с последним слоем (классификатор), а веса базовой сети для извлечения признаков оставались фиксированными.

Для проведения экспериментов был использован ПК с Nvidia GeForce GTX 1080 Ti GPU (12 Гбайт), AMD Ryzen Threadripper 1920X ЦПУ (2.2 ГГц), 64 Гбайт ОЗУ. В табл. 1 и 2 приведены результаты для разных архитектур и двух вариантов многозадачной СНС.

Из сравнения результатов видно, что несмотря на одинаковое число параметров и время классификации (табл. 2), модели с общим выходом (см. рис. 1) в среднем оказались на 1...3 % менее точными по сравнению с моделями с двумя классификаторами (см. рис. 2). Наличие нескольких выходов может считаться встроенной в архитектуру регуляризацией, позволяющей исключить влияние объектов одного вида на выходы, которые ответственны за другие подкатегории [2]. В результате подтверждается вывод о более высоком качестве многозадачных нейронных сетей [15, 16]: выходные слои (см. рис. 2) для собак и кошек обучаются отдельно, что позволяет более качественно настроить параметры для миноритарных классов (в данном случае, пород кошек) и, как следствие, понизить вероятность их ошибочной классификации. Наилучшую точность (90,6 %) показала архитектура Inception v3. При этом "легковесная" MobileNet v2 ($\alpha = 1{,}4$) показывает приемлемую точность, сравнимую с традиционной ResNet-50. По результатам качественного визуального анализа результатов классификации замечено, что ошибки допускаются либо при плохом качестве объекта-животного на изображении, либо для похожих пород.

Для сравнения в табл. 3 приведены известные наилучшие результаты для распознавания пород собак из набора данных Stanford Dogs. Как видно, предложенный подход с многозадачной СНС Inception оказывается на 0,6 %

Таблица 1

Точность классификации пород животных для многозадачных сверточных нейронных сетей

Базовая СНС	СНС с объединением всех подкатегорий		СНС с $N = 2$ выходными слоями	
	Собаки	Кошки	Собаки	Кошки
Inception ResNet v2	0,899	0,815	0,9	0,874
ResNet-50	0,869	0,809	0,859	0,879
ResNet-101	0,872	0,855	0,878	0,884
Inception v3	0,911	0,8	0,906	0,865
MobileNet v2 ($\alpha = 1.0$)	0,788	0,755	0,818	0,84
MobileNet v2 ($\alpha = 1.4$)	0,832	0,844	0,851	0,883

Таблица 2

Размер модели и среднее время классификации одного изображения

Базовая СНС	Число весов, млн	Время классификации, мс
Inception ResNet v2	54,75	10,1
ResNet-50	23,87	6,8
ResNet-101	43,00	9,8
Inception v3	22,55	9,0
MobileNet v2 ($\alpha = 1.0$)	2,48	6,1
MobileNet v2 ($\alpha = 1.4$)	4,62	6,9

Таблица 3

Результаты наилучших известных методов классификации пород собак

Модель	Число изображений каждой породы в обучающем множестве	Точность
Inception ResNet v2 [17]	100	0,900
ResNet-101 [17]	100	0,869
Inception v3 [17]	100	0,889
ResNet50 [18]	100	0,838
ResNet50-CURL [18]	100	0,816
MobileNetV2 [18]	100	0,789
MobileNetV2-CURL [18]	100	0,747
Вероятностная нейронная сеть с проекционными оценками [19, 20]	10	0,729

точнее по сравнению с лучшим известным методом [15].

Эксперимент 2: кластеризация фотографий домашних животных. **В** качестве материала для сравнения алгоритмов кластеризации был собран специальный набор [21] из 190 фотографий двух кошек черного и рыжего цвета (около 40 изображений каждой) и трех собак (рис. 6, см. третью сторону обложки). Большая часть фотографий собак принадлежит одной собаке (колли), при этом около половины ее фотографий сделаны в значительно младшем возрасте, поэтому эти фотографии изначально размечены как две отдельных собаки. Третья собака — черный лабрадор — присутствует примерно на 10 фотографиях. Кроме того, встречаются другие собаки, которые помещены в отдельный кластер выбросов.

В табл. 4 приведены число выделенных кластеров K и значения метрик оценки качества кластеризации в сравнении с реальным распределением по кластерам: Adjusted Rand Index (ARI) и Adjusted Mutual Information (AMI). Использовалась реализация методов кластеризации из библиотек scikit-learn и HDBSCAN. Указаны наилучшие комбинации параметров, вид животного (кошки и собаки группировались отдельно), а также используемый вектор признаков для кластеризации. В параметрах иерархической кластеризации приведен тип межкластерного расстояния, при этом во всех случаях наилучшее качество группировки достигалось для метрики L_1.

Здесь для кошек оптимальное число кластеров — два, а для собак корректными можно считать значения от 3 до 6. Значения ARI, AMI равны 1 при идеальной кластеризации, значения 0,5...0,6 указывают на получение приблизительно верных кластеров. Таким образом, в результате проведенных экспериментов было показано, что кластеризация вектора признаков **x**, извлеченных базовой СНС, может группировать фотографии домашних животных. В то же время для практического применения необходим тщательный выбор параметров для большого обучающего множества.

Заключение

В целом можно сделать заключение, что предложенный подход позволяет осуществить

Таблица 4

Сравнительный анализ методов кластеризации животных

Вид животного	Метод	Параметры	K	ARI	AMI
Собаки	Иерархическая кластеризация	Вектор признаков **x**, Ward	4	0,64	0,55
		Выходы СНС **p**, Average linkage	4	0,641	0,45
	DBSCAN	Вектор признаков **x**, eps = 9, core = 3	4	0,696	0,546
		Выходы СНС **p**, eps = 0.6, core = 3	4	0,549	0,418
	HDBSCAN	Вектор признаков **x**, minPts = 3	5	0,56	0,56
Кошки	Иерархическая кластеризация	Выходы СНС **p**, Complete linkage	2	0,9	0,845
	DBSCAN	Вектор признаков **x**, eps = 9, core = 4	2	1	1
		Выходы СНС **p**, eps = 0.5, core = 5	2	1	1
	HDBSCAN	Вектор признаков **x**, minPts = 3	2	1	1

высокоточное детектирование категорий нескольких различных видов объектов, для которых в обучающем множестве отсутствуют данные о положении на фотографии (обрамляющие прямоугольники). Показано, что для снижения затрат памяти можно использовать многозадачные нейронные сети (см. рис. 2) с несколькими выходами. Экспериментально показано, что такой подход позволяет повысить точность классификации подклассов (пород) домашних животных по сравнению с известными аналогами на основе специализированных нейронных сетей (см. табл. 3). Показано, что обученная нами многозадачная сеть позволяет извлекать характерные признаки объектов, приемлемые для группировки фотографий, содержащих одинаковые объекты (табл. 4).

Основным ограничением предлагаемого подхода является использование идентичных характерных признаков для разных видов объектов. Если для некоторых видов (например, изображений животных) такой подход является приемлемым, то для существенно различающихся объектов наилучшая точность достигается с использованием собственных специализированных СНС. Поэтому в будущих исследованиях необходимо модифицировать многозадачную СНС так, чтобы извлекать характерные признаки на нескольких различных слоях. Выходы первых слоев обычно в достаточной степени независимы от предметной области, поэтому могут быть использованы для классификации совершенно разных видов объектов, но при этом требуют больших объемов обучающих данных. Классификаторы выходов последних слоев могут обучаться даже на малых выборках наблюдений за счет использования доменной адаптации и технологии переноса знаний [2].

Список литературы

1. **Гречихин И. С., Савченко А. В.** Метод анализа предпочтений пользователя по фото- и видеоизображениям на мобильном устройстве на основе нейросетевых детекторов объектов на изображениях // Информационные технологии. 2019. Т. 25. № 9. С. 538—544

2. **Goodfellow I., Bengio Y., Courville A.** Deep Learning (Adaptive Computation and Machine Learning series) // Cambridge, USA, MIT Press, 2016. 800 p.

3. **Ren S., He K., Girshick R., Sun J.** Faster R-CNN: Towards real-time object detection with region proposal networks // Advances in neural information processing systems (NIPS). 2015. P. 91—99.

4. **Savchenko A. V.** Efficient facial representations for age, gender and identity recognition in organizing photo albums using multi-output ConvNet // PeerJ Computer Science. 2019. Vol. 5, p. 197.

5. **Khosla A., Jayadevaprakash N., Yao B., Li F.-F.** Novel dataset for fine-grained image categorization: Stanford dogs // Proceedings of the CVPR Workshop on Fine-Grained Visual Categorization (FGVC). 2011. Vol. 2.

6. **Parkhi O., Vedaldi A., Zisserman A., Jawahar C.** Cats and dogs // Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR). IEEE. 2012. P. 3498—3505.

7. **Grechikhin I., Savchenko A. V.** User modeling on mobile device based on facial clustering and object detection in photos and videos // Proceedings of the Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA). Springer. 2019. P. 429—440.

8. **Sandler M., Howard A., Zhu M., Zhmoginov A., Chen L. C.** MobilenetV2: Inverted residuals and linear bottlenecks // Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR). IEEE. 2018. P. 4510—4520.

9. **Lin T. Y., Goyal P., Girshick R., He K., Dollár P.** Focal loss for dense object detection // Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR). IEEE. 2017. P. 2980—2988.

10. **Szegedy C., Vanhoucke V., Ioffe S., Shlens J., Wojna Z.** Rethinking the Inception architecture for computer vision // Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR). IEEE. 2016. P. 2818—2826.

11. **Kolesnikov A., Beyer L., Zhai X., Puigcerver J., Yung J., Gelly S., Houlsby N.** Big Transfer (BiT): General Visual Representation Learning // arXiv preprint arXiv: 1912.11370. 2019 (дата доступа 29.04.2020).

12. **Theodoridis S., Koutroumbas K.** Pattern Recognition, 4th Ed. 2009. 984 p.

13. **McInnes L., Healy J., Astels S.** hdbscan: Hierarchical density based clustering // Journal of Open Source Software. 2017. Vol. 2. N. 11. doi:10.21105/joss.00205.

14. **Разработанное** Android-приложение. URL: https://drive.google.com/open?id=1rThhcKReOb5A9LBIH6jkP8tTiYjoVNWH (дата доступа 29.04.2020).

15. **Liu S., Johns E., Davison A. J.** End-to-end multi-task learning with attention // Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR). IEEE. 2019. P. 1871—1880.

16. **Yan C., Zhou L., Wan Y.** A multi-task learning model for better representation of clothing images // IEEE Access. 2019. Vol. 7. P. 34499—34507.

17. **Eshratifar A. E., Eigen D., Gormish M., Pedram M.** Coarse2Fine: a two-stage training method for fine-grained visual classification // arXiv preprint arXiv: 1909.02680. 2019 (дата доступа 29.04.2020).

18. **Luo J.-H., Wu J.** Neural network pruning with residual-connections and limited-data // arXiv preprint arXiv:1911.08114. 2019 (дата доступа 29.04.2020).

19. **Savchenko A. V.** Probabilistic neural network with complex exponential activation functions in image recognition // IEEE Transactions on Neural Networks and Learning Systems. 2020. Vol. 31, Iss. 2. P. 651—660

20. **Савченко А. В.** Тригонометрическая система функций в проекционных оценках плотности вероятности нейросетевых признаков изображений // Компьютерная оптика. 2018. Т. 42, № 1. С. 149—158.

21. **Набор** изображений для тестирования кластеризации домашних животных, URL: https://drive.google.com/drive/folders/1-tNB_GR2LkCBsNKQkxB—9ertdllgD7h (дата доступа 29.04.2020)

A. V. Savchenko, Doctor of Sciences, Professor, e-mail: avsavchenko@hse.ru,
I. S. Grechikhin, Postgraduate Student, Senior Lecturer, e-mail: igrechikhin@hse.ru,
National Research University Higher School of Economics, Nizhny Novgorod, Russian Federation

Detection of Specialized Object Categories in Photos from Mobile Device Based on a Multi-Task Neural Network

In this paper we consider the task of user preferences analysis for recommender engines based on a gallery of his or her mobile device. In particular, we propose the novel three-phase method for simultaneous image-based detection and recognition of particular objects. Conventional object detection techniques cannot be applied if there are many categories of the same object (pet breeds, car models, etc.) and there is a lack of large dataset with known bounding boxes for each object category. In order to deal with this issue, we estimate the borders of base objects (dogs, cats, cars, etc.) by using such existing neural network architectures as high precision Faster R-CNN or fast single-shot detectors. Secondly, the visual features (embeddings) of each object are extracted by using a multi-task convolutional neural network model with several outputs — one for each type of object. Finally, these embeddings are used to predict the concrete categories and group different photos of the same object by using cluster analysis techniques. The proposed approach is implemented in a special mobile application for Android. Experimental results for recognizing dog and cat breeds are presented. It is demonstrated that our method makes it possible to improve the accuracy of dog detection and recognition when compared to the known single-task neural nets. Moreover, we gather a special dataset of real photos with pets to estimate the clustering quality. It is shown that the L_1-normed features extracted by our multi-task model may be grouped rather accurately if hierarchical agglomerative clustering or HDBSCAN method are used.

Keywords: *image processing, convolutional neural networks, mobile systems, pet breed recognition, hierarchical clustering, multi-task learning, object detection*

Acknowledgments. The article was prepared within the framework of the Academic Fund Program at the National Research University Higher School of Economics (HSE University) in 2019-2020 (grant No. 19-04-004) and by the Russian Academic Excellence Project "5-100"

DOI: 10.17587/it.26.586-593

References

1. **Grechikhin I., Savchenko A. V.** Analysis of user preferences using photos and videos from mobile device based on object detection and neural networks, *Informacionnye tehnologii*, 2019, vol. 25, no. 9, pp. 538—544 (in Russian).

2. **Goodfellow I., Bengio Y., Courville A.** Deep Learning (Adaptive Computation and Machine Learning series). Cambridge, USA, MIT Press, 2016. 800 p.

3. **Ren S., He K., Girshick R., Sun** J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, *Advances in neural information processing systems (NIPS)*, 2015, pp. 91—99.

4. **Savchenko A. V.** Efficient facial representations for age, gender and identity recognition in organizing photo albums using multi-output ConvNet, *PeerJ Computer Science*, 2019, 5:e197.

5. **Khosla A., Jayadevaprakash N., Yao B., Li F.-F.** Novel dataset for fine-grained image categorization: Stanford dogs, *Proceedings of the CVPR Workshop on Fine-Grained Visual Categorization (FGVC)*, 2011, vol. 2.

6. **Parkhi O., Vedaldi A., Zisserman A., Jawahar C.** Cats and dogs, *Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2012, pp. 3498—3505.

7. **Grechikhin I., Savchenko A. V.** User modeling on mobile device based on facial clustering and object detection in photos and videos, *Proceedings of the Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA)*, Springer, 2019, pp. 429—440.

8. **Sandler M., Howard A., Zhu M., Zhmoginov A., Chen L. C.** MobilenetV2: Inverted residuals and linear bottlenecks, *Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2018, pp. 4510—4520.

9. **Lin T. Y., Goyal P., Girshick R., He K., Dollár P.** Focal loss for dense object detection, *Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2017, pp. 2980—2988.

10. **Szegedy C., Vanhoucke V., Ioffe S., Shlens J., Wojna Z.** Rethinking the Inception architecture for computer vision, *Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2016, pp. 2818—2826.

11. **Kolesnikov A., Beyer L., Zhai X., Puigcerver J., Yung J., Gelly S., Houlsby N.** Big Transfer (BiT): General Visual Representation Learning, arXiv preprint arXiv: 1912.11370. 2019 (date of access 29.04.2020).

12. **Theodoridis S., Koutroumbas K.** *Pattern Recognition*, 4th Edition, 2009, 984 p.

13. **McInnes L., Healy J., Astels S.** hdbscan: Hierarchical density based clustering, *Journal of Open Source Software*, 2017, vol. 2, no. 11, doi:10.21105/joss.00205.

14. Developed Android-application, available at: https://drive.google.com/open?id=1rThhcKReOb5A9LBIH6jkP8tTiYjoVNWH (date of access 29.04.2020)

15. **Liu S., Johns E., Davison A. J.** End-to-end multi-task learning with attention, *Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2019, pp. 1871—1880.

16. **Yan C., Zhou L., Wan Y.** A multi-task learning model for better representation of clothing images, *IEEE Access*, 2019, vol. 7, pp. 34499—34507.

17. **Eshratifar A. E., Eigen D., Gormish M., Pedram M.** Coarse2Fine: a two-stage training method for fine-grained visual classification, arXiv preprint arXiv: 1909.02680. 2019 (date of access 29.04.2020).

18. **Luo J.-H., Wu J.** Neural network pruning with residual-connections and limited-data, arXiv preprint arXiv:1911.08114. 2019 (date of access 29.04.2020).

19. **Savchenko A. V.** Probabilistic neural network with complex exponential activation functions in image recognition, *IEEE Transactions on Neural Networks and Learning Systems*, 2020, vol. 31, iss. 2, pp. 651—660.

20. **Savchenko A. V.** Trigonometric series in orthogonal expansions for density estimates of deep image features, *Computer Optics*, 2018, vol. 42, no. 1, pp. 149-158 (in Russian).

21. **Dataset** of cats and dogs images for testing of clustering, available at: https://drive.google.com/drive/folders/1-tNB_GR2LkCBsNKQkxB—9ertdllgD7h (date of access 29.04.2020).

УДК 621.311.243 DOI: 10.17587/it.26.594-601

С. В. Валевич, аспирант,
В. С. Осипович, канд. техн. наук, доц., e-mail: v.osipovich@bsuir.by,
Белорусский государственный университет информатики и радиоэлектроники, г. Минск, Беларусь,
И. Крузе, директор**,** e-mail: ingmar.kruse@sunsniffer.de,
"Санснифер" ООО, Ньюрнберг, Германия,
Р. М. Асимов, канд. техн. наук, директор, e-mail: roustam.asimov@sensotronica.com,
"Сенсотроника" ООО, г. Минск, Беларусь

Информационное обеспечение мониторинга технического состояния солнечных электростанций

Представлены результаты апробации программного средства, реализующего концепцию цифрового двойника для солнечной панели. Продемонстрирована возможность перехода от временных рядов измеренных параметров солнечной панели к вектору внутренних электрических параметров, полученных в результате создания ее цифрового двойника. Показана возможность автоматизации детектирования неисправных солнечных панелей. Проанализирована взаимосвязь видов неисправностей солнечных панелей и точности работы цифрового двойника.

Ключевые слова: *солнечная панель, цифровой двойник, физико-математическая модель, вольт-амперная характеристика, временной ряд параметров, внутренние электрические параметры модели*

Введение

В настоящее время в мире функционируют солнечные электростанции с установленной мощностью более 500 ГВт. Актуальной для таких источников энергии является проблема быстрого и своевременного устранения неисправностей, связанных с функционированием солнечных панелей. Эта проблема актуальна как для крупных солнечных электростанций (более одного ГВт), так и для домашних электростанций (3...5 кВт). Около 2 % солнечных панелей выходят из строя или теряют более 20 % своей эффективности в течение 11...12 лет работы [1]. Загрязнение солнечных панелей пылью тоже приводит к существенным потерям производительности солнечной электростанции [2].

Одним из способов определения неисправностей в солнечных панелях являются периодические термографические исследования [3—7]. Недостатками такого способа являются догорогостоящее оборудование, затраты времени на исследования, необходимость дополнительных инструментов для анализа потерь из-за загрязненности панелей. Кроме того, термографические исследования не позволяют определять отклонения в значениях шунтирующих диодов солнечных панелей [8].

Другой путь — использование систем сбора и анализа телеметрических данных [9—13]. Анализируя по каждой солнечной панели телеметрическую информацию, снятую специальной аппаратурой [14, 15], можно определять наименее эффективные панели и выяснять причины такой их работы. Кроме того, в этом случае могут быть использованы методы повышения эффективности работы солнечной электростанции [16—18]. Для поиска неисправных панелей необходим скрупулезный анализ телеметрической информации в виде временных рядов по каждой из них и последующая проверка выявленных панелей в лаборатории.

В связи с этим более перспективным и выгодным во всех отношениях видится использование цифровых двойников солнечных панелей для анализа текущего состояния электростанции и прогнозирования ее работы в будущем. Авторы работы [19] изложили концепцию использования физико-математической модели фотоэлектрической ячейки [20—25] для углубленного анализа работы, поиска неисправностей, прогнозирования работоспособности солнечных

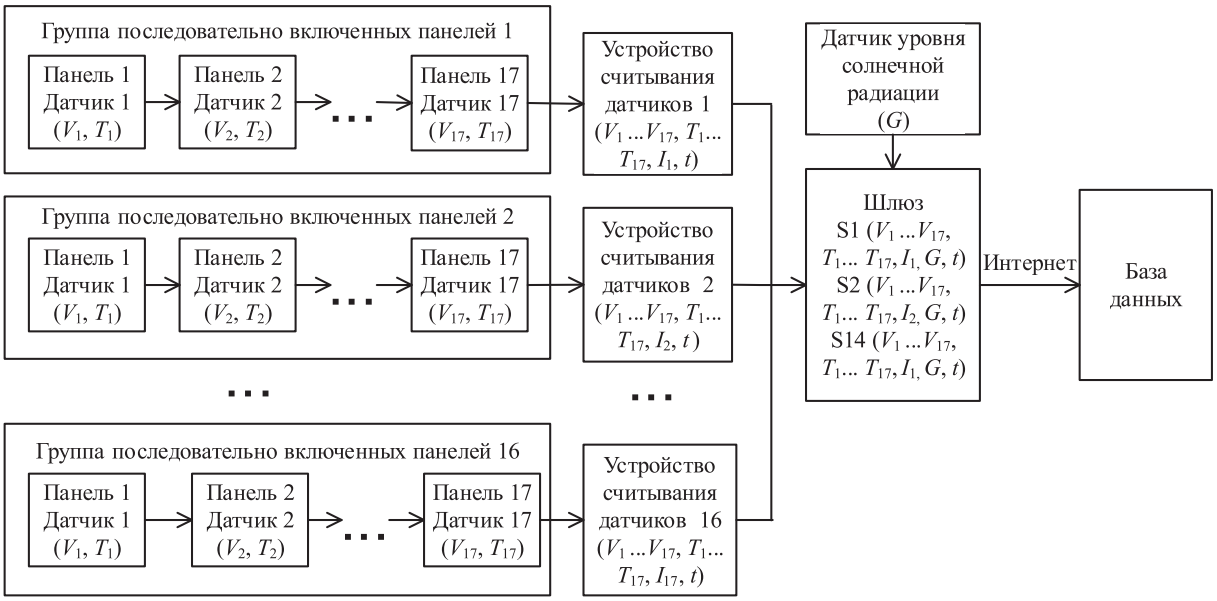


Рис. 1. Структурная схема солнечной электростанции и ситемы сбора телеметрической информации

элетростанций и осуществили первичные исследования по апробации этой концепции.

Целью настоящей работы явилось исследование возможности использования цифрового двойника солнечной панели для оценки технического состояния солнечной электростанции.

Исходные данные и методика эксперимента

Для проверки гипотезы были использованы результаты телеметрии солнечной электростанции с установленной мощностью 45,2 кВт, оборудованную 272 соленчными панелями M190 (STORM Energy GmbH, Germany). Структурная схема включения солнечных панелей и системы сбора телеметрической информации отражена на рис. 1.

Напряжение V и температура T в каждой солнечной панели измеряются каждые 15 минут с фиксацией даты и времени t посредством специального датчика Sensor (SunSniffer GmbH & Co, Germany). Данные с группы семнадцати последовательно включенных панелей собираются в устройстве считывания датчиков — String Reader (SunSniffer GmbH & Co, Germany), который дополнительно измеряет силу тока I в цепи этих панелей. Датчик уровня энергетической экспозиции Irradiance Sensor Si-13TC (Ingenieurbüro Mencke & Tegtmeyer, Germany) обеспечивает измерение ее значения G и его передачу в шлюз — Gateway (SunSniffer GmbH & Co), который в свою очередь записывает все результаты телеметрии в базу данных на сервер. В эксперименте были использованы 20160 векторов (t, I, V, T, G), собранных в течение 7 месяцев работы солнечной электростанции с мая по ноябрь 2019 г. Объем информации о векторах (t, I, V, T, G) составил 269 Мбайт для всего времени наблюдений.

Для проверки деффектных солнечных панелей в лабораторных условиях использовали установку Spi-Sun Simulator 4600 SLP (Spire Solar, The Netherlands). В ходе испытаний были сняты вольт-амперные характеристики при стандартных условиях (flash of STC).

В качестве физико-математической модели для построения цифрового двойника солнечной панели была использована двухдиодная модель солнечной ячейки [23, 24]. Эквивалентная схема двухдиодной модели отражена на рис. 2.

Сила тока, вырабатываемая солнечной панелью, согласно двухдиодной физико-математическая модели определяется уравнением

$$I = I_{ph} - I_{0_1}\left(\mathbf{e}^{\frac{V+IR_s}{n_{s1}v_T}} - 1\right) - I_{0_2}\left(\mathbf{e}^{\frac{V+IR_s}{n_{s2}v_T}} - 1\right) - \frac{V + IR_s}{R_p}, \quad (1)$$

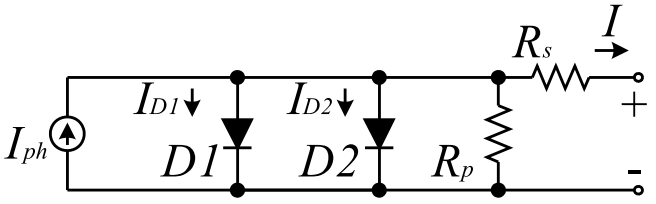


Рис. 2. Схема электрическая принципиальная двухдиодной модели фотоэлектрической ячейки

где I_{ph} — фототок (А); I_{0_1} — обратный ток насыщения диода D1 (А), характеризует силу тока рекомбинации носителей заряда в базе, эмиттере и подложке; I_{0_2} — обратный ток насыщения диода D2 (А), характеризует силу тока рекомбинации носителей заряда в области p-n перехода; n_{s1} — коэффициент идеальности диода D1; n_{s2} — коэффициент идеальности диода D2; v_T — термоЭДС (В); R_s — последовательное сопротивлением (Ом); R_p — шунтирующее сопростивление (Ом); I — сила тока в солнечной панели (А); V — напряжение в солнечной панели (В).

Для построения цифрового двойника уравнение (1) было преобразовано к следующему виду:

$$V = v_T \ln\left[\mathbf{e}^{-\frac{V+IR_s}{R_p}}\left(\left(\frac{I_{ph} - I_{0_1} + I_{0_2} - I}{I_{0_2}}\right) - \frac{V + IR_s}{R_p I_{0_2}}\right) - 1\right] IR_s; \quad (2)$$

$$v_T = \frac{v_{T1} v_{T2}}{v_{T1} - v_{T2}}.$$

ТермоЭДС v_T для каждого диода в отдельности определяется выражением

$$v_{T1(2)} = n_{s1(2)} kT/q,$$

где k — постоянная Больцмана ($1{,}3865 \cdot 10^{-23}$ Дж/К); T — температура в солнечной панели (К); q — заряд электрона ($1{,}6021 \cdot 10^{-19}$ С).

Зависимость фототока I_{ph} от уровня солнечной радиации G описывается выражением

$$I_{ph} = \frac{G}{G_{STC}}((I_{sc})_{T_1} + K_0(T - T_1)), \quad (3)$$

где I_{sc} — сила тока короткого замыкания, составляет 5,44 А для модели солнечных панелей, задействованных в эксперименте; G — энергетическая экспозиция (Вт/м^2); G_{STC} — энергетическая экспозиция при нормальных условиях проведения испытаний (1000 Вт/м^2); K_0 — температурный коэффициент (0,033 %/К); T — рабочая температура (К); T_1 — нормальная температура (293,15 К).

При решении нелинейного уравнения (2) приняты следующие дополнительные допущения. В реальных фотодиодах значения обратных токов насыщения I_{0_1} и I_{0_2} отличаются на порядок. Поиск этих значений при решении уравнения (2) значительно усложняется при незначительной потере в точности расчетов. Поэтому обратный ток насыщения диода I_0 в целях упрощения был рассчитан по формуле [22]

$$I_0 = I_{0_1} = I_{0_2} = (I_0)_{T_1}\left(\frac{T}{T_1}\right)^3 \mathbf{e}^{\frac{q(E_g)_{T_1}}{nk}\left(\frac{1}{T_1} - \frac{1}{T}\right)}, \quad (4)$$

$$(I_0)_{T_1} = \left(\frac{(I_{sc})_{T_1}}{\mathbf{e}^{\frac{q(V_{OC})_{T_1}}{nkT_1}} - 1}\right), \quad (5)$$

где I_{SC} — сила тока короткого замыкания; E_g — ширина запрещенной зоны полупроводника; V_{OC} — напряжение холостого хода. Обратный ток насыщения был расчитан отдельно для каждого диода физико-математической модели [22].

Разработанное программное средство работает следующим образом:

1. Результаты телеметрии преобразуются в векторы входных данных (t, G, I, V, T). Преобразование необходимо в связи с тем, что результаты телеметрии не структурированы: не имеют привязки к единой временной точке.

2. Векторы входных данных (t, G, I, V, T) проходят через фильтр. Отсеиваются точки, не соответствующие следующим условиям: сила тока — 0...15 А, уровень энергетической экспозиции — 360...1500 Вт/м^2, производная по току −0,2...+0,2, производная по уровню энергетической экспозиции −3,8...+3,8.

3. При наличии пятидесяти векторов осуществляется решение нелинейного уравнения (2) методом наименьших квадратов, совмещенным с генетическим алгоритмом, для каждого параметра: последовательное и шунтирующее сопростивления, коэффициенты идеальности диодов, обратный ток насыщения диодов. Результатом решения уравнения является вектор внутренних параметров физико-математической модели (R_p, R_s, n_1, n_2, I_0).

4. Вектор внутренних парамеров подвергается анализу, используется для расчета потерь энергии в результате затенения, запыления или наличия неисправности в солнечной панели.

5. Вектор внутренних параметров используется для расчета силы тока и напряжения в солнечной панели при заданой температуре и уровне солнечной радиации, измеренных датчиками.

Для проверки точности расчетов был использован закон Кирхгофа, а именно:

$$I_{ph} - I_{out} - I_{loss} = 0, \quad (6)$$

где I_{ph} — сила тока, индуцированная солнечным излучением; I_{out} — сила тока, измеренная

на выходе солнечной панели; I_{loss} — сила тока, характеризующая потери в солнечной панели.

Однако в случае подстановки в выражение (6) неверных значений результат вычислений не будет равен нулю. Этот факт был использован для проверки значений, рассчитанных с помощью цифрового двойника. Аналитическая погрешность ΔI цифрового двойника солнечной панели находится из уравнения (6):

$$I_{ph} - I_{out} - I_{loss} = \Delta I. \quad (7)$$

Результаты и их обсуждение

Анализ временных рядов напряжений и силы тока (рис. 3, см. четвертую сторону обложки) показывает, что выявление дефектных солнечных панелей затруднено и требует последовательного сравнения временных рядов напряжений или выработанной мощности заведомо исправной панели. Кроме того, этот анализ результатов телеметрии необходимо осуществлять по временным рядам, снятым в ясные солнечные дни.

В результате использования разработанного программного средства получаем вектор внутренних параметров физико-математической модели (R_p, R_s, n_1, n_2, I_0). Для большей наглядности переносим значения R_p, R_s и I_0, всех панелей электростанции в трехмерное пространство (рис. 4). Такое представление позволяет быстро выделить из всех 238 панелей электростанции две с аномальными значениями всех трех внутренних параметров физико-математической модели (справа сверху на рис. 4). Номера этих солнечных панелей 16_11 и 7_17.

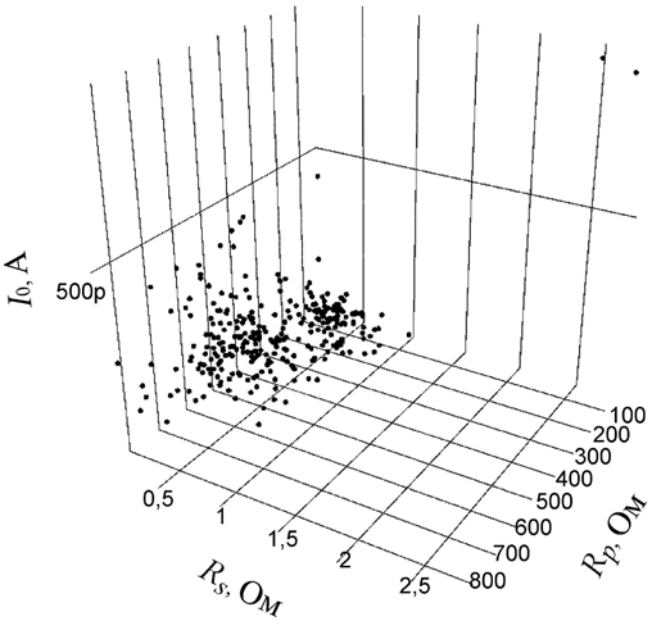


Рис. 4. Результат построения векторов значений (обратный ток насыщения диода I_0, последовательное сопротивление R_s, шунтирующее сопростивление R_p) всех панелей солнечной электростанции в трехмерном пространстве

В таблице приведены результаты решения уравнения (2) для пяти выбранных случайным образом солнечных панелей и для панелей с номерами 16_11 и 7_17.

Как видно из таблицы, солнечные панели 16_11 и 7_17 имеют низкое значение шунтирующего сопротивления R_p. Это может быть причиной увеличенного значении внутренних потерь энергии.

Результат расчета внутренних электрических параметров некоторых солнечных панелей

Номер солнечной панели	Внутренние параметры модели			
	R_p, Ом	R_s, Ом	ns_1/ns_2	I_0, A·10^{-10}
4_12	718	0,604	1/1	2,08
6_10	754	0,564	1/1	1,74
8_15	715	0,505	1/1	1,52
13_7	625	0.683	1/1	1.82
14_16	645	0,4258	1/1	1,5
16_11	85	3,0215	1/1	1,42
7_17	163	2,613	1/1	1,58

Сравнение (рис. 5, см. четвертую сторону обложки) фактических значений уровня напряжения солнечной панели 16_11 на временных рядах с другими показывает, что на ней имеются падения напряжения в утреннее и вечернее время.

Причиной такого поведения может быть влияние двух факторов. Во-первых, снижение значения шунтирующего сопротивдения R_p приводит к росту силы тока через это сопротивление. Используя закон Ома и значение напряжения в оптимальной точке V_{MPP} (для имеющихся панелей оно составляет 36,5 В), получим силу тока 0,43 А. Такая сила тока окажет влияние на вольт-амперную характеристику солнечной панели.

Во-вторых, потери тока на шунтирущем сопротивлении складываются с потерями тока в оптимальной точке работы инвертора (рис. 6, см. четвертую сторону обложки).

При более низких значениях солнечного излучения кривая вольт-амперной характери-

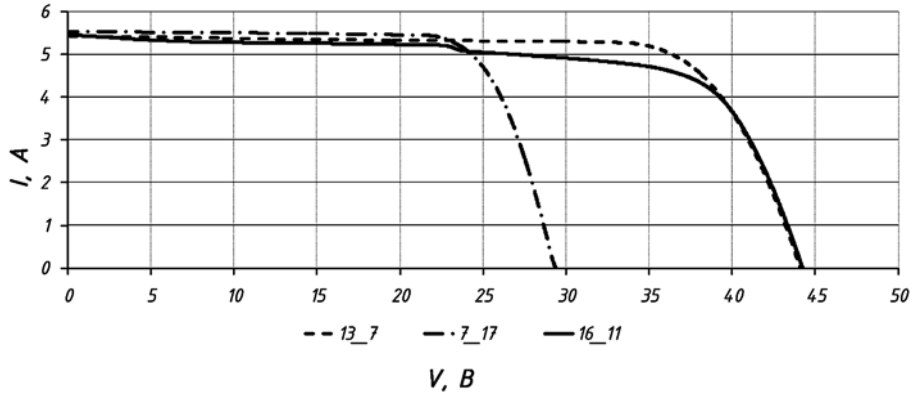


Рис. 7. Вольтамперные характеристики, измеренные при стандартных условиях (1000 В/м^2, 25 °С) для солнечных панелей: 13_7, 7_17 и 16_11

стики (ВАХ) смещается в сторону уменьшения силы тока. На рис. 6 показано, что одинаковый прирост силы тока на ВАХ для низких значений уровня энергетической экспозиции приведет к заметно большему падению напряжения по сравнению с более высокими значениями энергетической экспозиции. Другими словами, увеличение тока за счет внутренних потерь приводит к большим падениям напряжения при более низких значениях энергетической экспозиции (утром и вечером) и очень небольшому изменению напряжения при более высоких значениях энергетической экспозиции. Этот вывод подтверждается поведением временного ряда напряжений этой солнечной панели, отраженным на рис. 5 (см. четвертую сторону обложки).

Для проверки наличия отклонений от нормальной работы солнечных панелей 7_17 и 16_11 в лабораторных условиях были осуществены их исследования при работе в стандартных условиях (1000 Вт/м^2, 25 °С). Результаты измерений показаны на рис. 7. Напряжение для солнечной панели 7_17 смещено в сторону снижения и составляет около 23 В. Это означает, что примерно треть солнечных элементов фотоэлектрического модуля не работает из-за активации защитного диода, что, в свою очередь, может быть вызвано разрывом цепи фотоэлементов. Снижение мощности, генерируемой фотоэлектрическим модулем 16_11, происходит из-за уменьшения сопротивления изоляции. Это вызвало ток утечки и смещение оптимальной точки, что коррелирует с предположениями, изложенными при обсуждении рис. 4 и таблицы.

Размещение в трехмерном пространстве вектора (сила тока в точке оптимальной мощности I_{MPP}, аналитическая погрешность по току I_s, мощность в оптимальной точке P_{MPP}), полученного с использованием данных телеметрии за август 2019 года, тоже позволяет визуализировать проблемные модули (рис. 8). Точки, обозначенные на рис. 8 цифрой *6,* — это результат построения в трехмерном пространстве векторов внутренних электрических параметров для панелей с номерами 16_11 и 7_17. Как видно из рис. 8, такое представление позволяет идентифицировать солнечные панели с дефектами.

Кроме того, векторы (I_{MPP}, ΔI, P_{MPP},) для всех панелей позволяют выделить дополнительно пять группировок панелей (рис. 8). Группировка панелей 1 характеризуется наибольшей силой тока в точке оптимальной мощности — это солнечные панели последовательных включений 13, 14 и 15. Векторы (I_{MPP}, ΔI, P_{MPP}) для этих панелей находятся в диапазонах (4,927...4,973 А, 0,026...0,035 А, 176...191 Вт). Группы последовательно включенных панелей, входящих в состав группировки 1, имеют наклон установки относительно поверхности Земли 25°, остальные панели электростанции — 21°. Оптимальным углом установки панелей для Ньюрнберга считается угол в 26° [26]. Следует отметить, что группа последовательно включенных панелей с номером 16 тоже имеет угол установки относительно Земли 25°. Однако панели этой группы попали в группировку 2 (рис. 8). Это связано с тем, что в со-

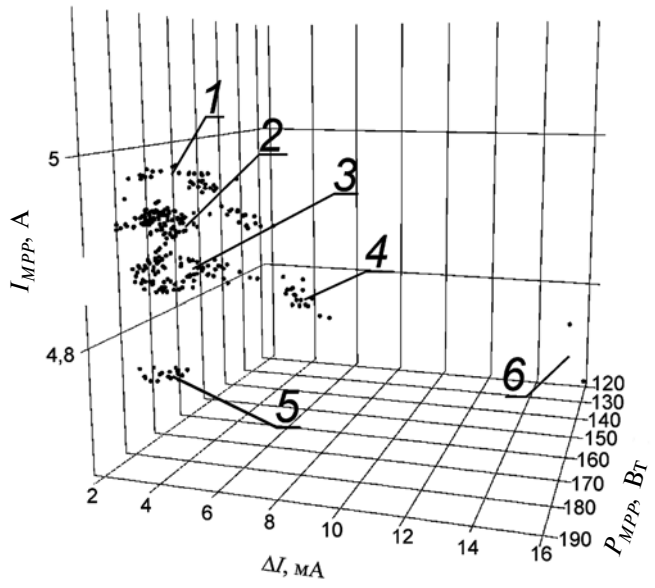


Рис. 8. Результат построения векторов значений (сила тока в точке оптимальной мощности I_{MPP}, аналитическая погрешность по току ΔI, мощность в оптимальной точке P_{MPP}) всех панелей солнечной электростанции в трехмерном пространстве

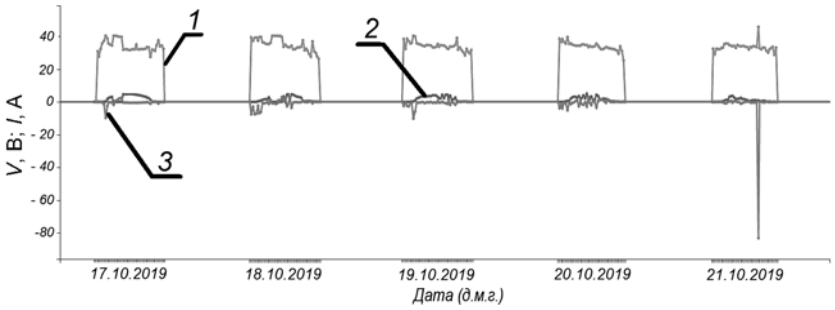


Рис. 9. Временные ряды напряжения (*1*), силы тока (*2*) и аналитической погрешности по току (*3*), солнечной панели 6_17

ставе этой группы находится панель 16_11. Она является дефектной и имеет низкую силу тока в оптимальной точке (4,78 А). Вся группа последовательно включенных панелей по этой причине также имеет более низкую силу тока в оптимальной точке — 4,913 А. Наличие в составе группы 16 дефектной панели привело к тому, что общий режим работы этой группы панелей стал подобен режиму работы групп панелей с углом установки 21°, а аналитическая погрешность составила 0,036... 0,048 А.

Компоненты векторов (I_{MPP}, ΔI, P_{MPP},) для панелей группировки 2 имеют следующие значения: 4,875...4,928 А, 0,016...0,038 А, 170...186 Вт. Значения аналитической погрешности у панелей этой группировки характеризуются более широким диапазоном в сравнении со значениями группировки 1. Это может быть связано с наличием панелей, имеющих частичное или более интенсивное запыление, а также панелей со сниженной эффективностью преобразования солнечной энергии по другим причинам.

Компоненты векторов (I_{MPP}, ΔI, P_{MPP},) для панелей группировки 3 имеют следующие значения: 4,81...4,874 А, 0,016...0,038 А, 165...181 Вт.

Группировка 4 имеет силу тока в оптимальной точке I_{MPP}, мощность в оптимальной точке $P_{MPP,}$ характерные для группировки 3. Однако аналитическая погрешность группировки 4 выше и составляет 0,041...0,059 А. В группировку 3 входит последовательно включенная группа панелей с номером 3. Анализ показал, что причина повышенной аналитической погрешности — это параллельное подключение групп 3 и 7 к одному входу инвертора. Группа 7 имеет в своем составе дефектную панель 7_17. Из-за этого группа 3 фактически работала в менее выгодной рабочей точке, чем она способна была работать. Это подтверждают рассчитанные по математической модели параметры оптимальной точки для панелей этой группы.

Наличие же в составе группы 7 дефектной панели стало причиной попадания его в группировку 4. Эта группировка характеризуется самыми низкими значениями силы тока в оптимальной точке I_{MPP}, мощности в оптимальной точке P_{MPP}. Вектор (I_{MPP}, ΔI, P_{MPP},) для этой группировки имеет значения (4,695...4,75 А, 0,022...0,028 А, 170...175 Вт).

Расчет аналитической погрешности для солнечных панелей электростанции показывает, что значение этого параметра, в основном, стремится к нулю. Однако ежедневно в промежутке с 10.15 до 10.30 аналитическая погрешность составляет значение от −8 до −12 А (рис. 9) для всех панелей электростанции. Это может быть связано с тем, что угол наклона солнечных панелей по отношению к положению солнца в это время суток способствует отражению значительной части солнечного излучения. В то же время защитное покрытие датчика солнечной радиации обеспечивает отсутствие отражения солнечного излучения. Расхождение в значениях измеренного и фактически достигшего фотодиодов уровня солнечной радиации приводит к большим значениям аналитической погрешности. Вместе с тем, это может быть связано с несовпадением фактических и измеренных значений напряжения и силы тока. Когда инвертор начинает потреблять ток, соответствующее значение напряжения приходит с задержкой, вызванной передачей данных по линии электропередачи от датчика.

Кроме того, анализ кривой аналитической погрешности позволяет определять противоречивую выборку данных. На рис. 9 пик напряжения в 13:45, 21.10.2019 г. не соответствует другим измеренным параметрам: силе тока, температуре и уровню солнечного излучения. Это приводит к пику на кривой аналитичекой погрешности.

Значение аналитической погрешности, стремящееся к нулю, указывает как на достоверную выборку данных, так и на точность решения уравнений физико-математической модели.

Выводы

Показана жизнеспособность концепции анализа телеметрических данных на солнечных электростанциях с использованием цифровых двойников солнечных панелей. Использованный в исследовании цифровой двойник основан на двухдиодной физико-математической модели, настраиваемой под конкретную солнечную

панель с помощью измеренных данных. Такой подход позволил нам осуществить переход от временных рядов данных, собранных в течение месяца для каждой солнечной панели, к векторам из семи внутренних электрических параметров: одна солнечная панель — один набор параметров в месяц. Размещение отдельных парметров в трехмерном пространстве дало возможность оценить состояние каждой солнечной панели электростанции и выявить две неисправные из 272 панелей.

Установлено, что неисправная солнечная панель снижает выходную мощность группы последовательно включенных панелей на 1,5...2 %. Кроме того, наличие неисправной панели снижает эфективность не только своей группы последовательно включенных панелей, но и группы панелей, подключенной к инвертору параллельно с ней.

Результаты исследований и разработанная технология найдут применение в автоматизации поиска неисправностей солнечных электростанций на основе систем мониторинга, а также при развитии подсистем поддержки принятия решений для систем мониторинга.

Список литературы

1. **Performance** and Reliability of Photovoltaic Systems. IEA International Energy Agency, External final report IEA-PVPS March 2014. P. 5.

2. **Mejia F., Kleissl J., Boscet J. L.** The effect of dust on solar photovoltaic systems // Energy Procedia. 2014. Vol. 49. P. 2375.

3. **Schuss C., Leppänen K., Remes K., Saarela J., Eichberger B., Fabritius T., Rahkonen T.** Detecting Defects in Photovoltaic Cells and Panels and Evaluating the Impact on Output Performances // IEEE Transactions on Instrumentation and Measurement. 2016. Vol. 65, N. 5. P. 1108—1119.

4. **Breitenstein O., Langenkamp M., Lang O., Schirrmacher A.** Shunts due to laser scribing of solar cells evaluated by highly sensitive lock-in thermography // Solar Energy Mater. Solar Cells. 2001. Vol. 65, N. 1. P. 55—62.

5. **Breitenstein O.** Nondestructive local analysis of current—voltage characteristics of solar cells by lock-in thermography // Solar Energy Mater. Solar Cells. 2011. Vol. 95, N. 10. P. 2933—2936.

6. **Ramspeck K., Bothe K., Hinken D., Fischer B., Schmidt J., Brendel R.** Recombination current and series resistance imaging of solar cells by combined luminescence and lock-in thermography // Appl. Phys. Lett. 2007. Vol. 90, N. 15. P. 153502.

7. **Fertig F., Greulich J., Rein S..** Spatially resolved determination of the short-circuit current density of silicon solar cells via lock-in thermography // Appl. Phys. Lett. 2014. Vol. 104, N. 20. P. 201111.

8. **Швец С. В., Байшев А. В.** Назначение шунтирующих диодов солнечной панели и методы их диагностики // Вестник Иркутского государственного технического университета. 2016. Т. 23(6). С. 1187—1202.

9. **List** of monitoring and control systems for photovoltaic solar panels generators. URL: https://photovoltaic-software.com/monitoring-control-solar-PV-inverter-non-dependent-(multi-brand).php.

10. **SunSniffer** / Whole plant at a glance? With SunSniffer's "Deep View" the functionality of each module is visible... 2017. URL: http://www.sunsniffer.de/solution/what-is-sunsniffer.html (Date of access: 3.12.2019).

11. **SolarEye.** URL: https://www.solareye.eu/platform/?r=site/page&view=features (Date of access: 5.06.2018).

12. **PVsyst.** URL: http://www.pvsyst.com/en/ (Date of access: 5.06.2018).

13. **PV_LIB.** URL:https://pvpmc.sandia.gov/applications/pv_lib-toolbox/ (Date of access: 5.06.2018).

14. **Method** for monitoring individual photovoltaic modules in an arrangement that comprises several photovoltaic modules and device for performing said: pat. US20120197569 international G01K13/00, G01R19/00, G01R27/00, G06F15/00 / Ingmar Kruse, Roustam Asimov; Ingmar Kruse — US 13/379,319 at 28.06.2010; public 2.08.2012 //United States Patent and Trademark Office. [Electronic resource] / USPTO Patent Full-Text and Image Database (WO 2011000505 A8, CN102473764A, CN102473764B, DE102009031839A1, DE102009031839B4, EP2449643A2, EP2449643B1, US9070281, US20120197569, WO2011000505A2, WO2011000505A3, WO2011000505A4).

15. **Method** for disconnecting a photovoltaic assembly and photovoltaic assembly: pat. US 20140311546 A1 international H01L31/18, H01L31/05 / Ingmar Kruse, Roustam Asimov; Ingmar Kruse — US 13/993,981 at 13.12.2011; public 23.10.2014 // United States Patent and Trademark Office [Electronic resource] / USPTO Patent Full-Text and Image Database (DE102010054354A1, EP2652857A1, WO2012079742A1).

16. **Нго С. К.** Повышение эффективности солнечных батарей с помощью следящей системы // Известия Тульского государственного университета. Технические науки. 2013. № 1. С. 318—321.

17. **Деменкова Т. А., Финенко А. А.** Аппаратная реализация алгоритмов для систем управления солнечными батареями // Вестник МГТУ МИРЭА. 2015. № 2. С. 7.

18. **Малинин Г. В., Серебрянников А. В.** Слежение за точкой максимальной мощности солнечной батареи // Вестник Чувашского университета. 2016. № 3. С. 76—92.

19. **Asimov R. M., Chernoshey S. V., Kruse I., Osipovich V. S.** Digital twin in the analysis of a big data // *BIG DATA Advanced Analytics*: collection of materials of the fourth international scientific and practical conference (Minsk, Belarus, May 3—4, 2018) / Editorial board: M. Batura [etc.]. Minsk, BSUIR, 2018. P. 68—77.

20. **Ishaque K., Salam Z., Taheri H., Syafaruddin.** Modeling and simulation of photovoltaic (PV) system during partial shading based on a two-diode model // Simulation Modelling Practice and Theory. 2011. N. 19. P. 1613—1626.

21. **Chang, Chih-Hao, Zhu J.-J., Tsai H.-L.** Model-based performance diagnosis for PV systems // SICE Annual Conference 2010, Proceedings of. IEEE, August 18-21, 2010, Taiwan. P. 2139—2145.

22. **Stegner C., Dalsass M., Luchscheider P., Brabec C. J.,** Monitoring and assessment of PV generation based on a combination of smart metering and thermographic measurement // Solar Energy. 2018. Vol. 163. P. 16—24.

23. **Hiren P., Agarwal V.** MATLAB-based modeling to study the effects of partial shading on PV array characteristics // IEEE Transactions On Energy Conversion. 2008. Vol. 23, N. 1. P. 302—310.

24. **Almonacid F., Rus C., Pérez-Higueras P., Hontoria L.** Calculation of the energy provided by a PV generator. Comparative study: Conventional methods vs. artificial neural networks // Energy. 2011. Vol. 36, Iss. 1. P. 375—384.

25. **Шарифов Б. Н., Терегулов Т. Р.** (2015). Моделирование солнечной панели в программе MATLAB/Simulink // Вестник Уфимского государственного авиационного технического университета. 2015. Т. 19, № 4 (70). С. 77—83.

26. **Solar** Angle Calculator. URL: http://www.solarelectricity-handbook.com/solar-angle-calculator.html (Date of access: 5.03.2019).

S. V. Valevich, Master of Engineering, Graduate Student, e-mail: whenthegroundcavedin@gmail.com,
V. S. Osipovich, PhD., Associate Professor, e-mail: v.osipovich@bsuir.by,
Belarussian State University of Informatics and Radioelectronics, Minsk, 220013, Republic of Belarus,
I. Kruse, SEO, e-mail: ingmar.kruse@sunsniffer.de,
SunSniffer GmbH & Co. KG, Nuremberg, 90489, Federal Republic of Germany,
R. M. Asimov, PhD., SEO, e-mail: roustam.asimov@sensotronica.com,
Sensotronica Ltd, Minsk, 220010, Republic of Belarus

Information Support for Monitoring of Solar Power Station's Technical State

Results of the software tool which implements the Digital Twin concept for the PV module are presented. The possibility of transition from the time series of the measured PV module parameters to the vector of internal electrical parameters obtained by creating its digital twin is demonstrated. The possibility of fault detection automation for PV modules is presented. The influence and interconnection of PV module fault types on the digital twin accuracy are analyzed.

Keywords: *PV module, Digital Twin, Physical and Mathematical Model, I-V Characteristics, Parameter Time Series, Internal Electrical Model Parameters*

DOI: 10.17587/it.26.594-601

References

1. **Performance** and Reliability of Photovoltaic Systems, IEA International Energy Agency, External final report IEA-PVPS, March 2014, 5 p.

2. **Mejia F., Kleissl J., Boscet J. L.** The effect of dust on solar photovoltaic systems, *Energy Procedia*, 2014, vol. 49, p. 2375.

3. **Schuss C., Leppänen K., Remes K., Saarela J., Eichberger B., Fabritius T., Rahkonen T.** Detecting Defects in Photovoltaic Cells and Panels and Evaluating the Impact on Output Performances, *IEEE Transactions on Instrumentation and Measurement,* 2016, vol. 65, no. 5, pp. 1108—1119.

4. **Breitenstein O., Langenkamp M., Lang O., Schirrmacher A.** Shunts due to laser scribing of solar cells evaluated by highly sensitive lock-in thermography, *Solar Energy Mater. Solar Cells.,* 2001, vol. 65, no. 1, pp. 55—62.

5. **Breitenstein O.** Nondestructive local analysis of current—voltage characteristics of solar cells by lock-in thermography, *Solar Energy Mater. Solar Cells.,* 2011, vol. 95, no. 10, pp. 2933—2936.

6. **Ramspeck K., Bothe K., Hinken D., Fischer B., Schmidt J., Brendel R.** Recombination current and series resistance imaging of solar cells by combined luminescence and lock-in thermography, *Appl. Phys. Lett.*, 2007, vol. 90, no. 15, pp. 153502.

7. **Fertig F., Greulich J., Rein S.** Spatially resolved determination of the short-circuit current density of silicon solar cells via lock-in thermography, *Appl. Phys. Lett.*, 2014, vol.104, no. 20, pp. 201111.

8. **Shvets S. V., Baishev A. V.** The purpose of the shunt diodes of the solar panel and methods for their diagnosis, *Vestnik Irkutskogo gosudarstvennogo tekhnicheskogo universiteta*, 2019, vol. 23, no. 6, pp. 1187—1202 (in Russian).

9. **List** of monitoring and control systems for photovoltaic solar panels generators, available at: https://photovoltaic-software.com/monitoring-control-solar-PV-inverter-non-dependent-(multi-brand).php

10. **SunSniffer,** Whole plant at a glance? With SunSniffer's "Deep View" the functionality of each module is visible... — 2017, available at: http://www.sunsniffer.de/solution/what-is-sunsniffer.html (Date of access: 3.12.2019).

11. **SolarEye**, available at: https://www.solareye.eu/platform/?r=site/page&view=features (Date of access: 5.06.2018).

12. **PVsyst**, available at: Mode of access http://www.pvsyst.com/en/ (Date of access: 5.06.2018).

13. **PV_LIB**, available at: Mode of access https://pvpmc.sandia.gov/applications/pv_lib-toolbox/ (Date of access: 5.06.2018).

14. **Method** for monitoring individual photovoltaic modules in an arrangement that comprises several photovoltaic modules and device for performing said: pat. US20120197569 international G01K13/00, G01R19/00, G01R27/00, G06F15/00 / Ingmar Kruse, Roustam Asimov; Ingmar Kruse — US 13/379,319 at 28.06.2010; public 2.08.2012 //United States Patent and Trademark Office [Electronic resource] / USPTO Patent Full-Text and Image Database (WO 2011000505 A8, CN102473764A, CN102473764B, DE102009031839A1, DE102009031839B4, EP2449643A2, EP2449643B1, US9070281, US20120197569, WO2011000505A2, WO2011000505A3, WO2011000505A4).

15. **Method** for disconnecting a photovoltaic assembly and photovoltaic assembly: pat. US 20140311546 A1 international H01L31/18, H01L31/05 / Ingmar Kruse, Roustam Asimov; Ingmar Kruse — US 13/993,981 at 13.12.2011; public 23.10.2014 // United States Patent and Trademark Office [Electronic resource] / USPTO Patent Full-Text and Image Database (DE102010054354A1, EP2652857A1, WO2012079742A1).

16. **Ngo S. K.** Improving the efficiency of solar panels with a tracking system, *Izvestiya Tul'skogo gosudarstvennogo universiteta. Tekhnicheskie nauki*, 2013, no. 1, pp. 318—321 (in Russian).

17. **Demenkova T. A., Finenko A. A.** Hardware implementation of algorithms for solar control systems, *Russian Technological Journal*, 2015, no. 2, pp. 7 (in Russian).

18. **Malinin G. V., Serebryannikov A. V.** Tracking the point of maximum power of the solar battery, *Vestnik CHuvashskogo universiteta*, 2016, no. 3 (in Russian).

19. **Asimov R. M., Chernoshey S. V., Kruse I., Osipovich V. S.** Digital twin in the analysis of a big data, *BIG DATA Advanced Analytics*: collection of materials of the fourth international scientific and practical conference (Minsk, Belarus, May 3—4, 2018) / Editorial board: M. Batura [etc.], Minsk, BSUIR, 2018, pp. 68—77.

20. **Ishaque K., Salam Z., Taheri H., Syafaruddin.** Modeling and simulation of photovoltaic (PV) system during partial shading based on a two-diode model, *Simulation Modelling Practice and Theory,* 2011, vol. 19, pp. 1613—1626.

21. **Chang, Chih-Hao, Zhu J.-J., Tsai H.-L.** Model-based performance diagnosis for PV systems, *SICE Annual Conference 2010, Proceedings of IEEE*, August 18—21, 2010, Taiwan, pp. 2139—2145.

22. **Stegner C., Dalsass M., Luchscheider P., Brabec C. J.,** Monitoring and assessment of PV generation based on a combination of smart metering and thermographic measurement, *Solar Energy*, 2018, vol. 163, pp. 16—24.

23. **Hiren P., Agarwal V.** MATLAB-based modeling to study the effects of partial shading on PV array characteristics, *IEEE Transactions On Energy Conversion,* 2008, vol. 23, no. 1, pp. 302—310.

24. **Almonacid F., Rus C., Pérez-Higueras P., Hontoria L.** Calculation of the energy provided by a PV generator. Comparative study: Conventional methods vs. artificial neural networks, *Energy*, 2011, vol. 36, iss.1, pp. 375—384.

25. **Sharifov B. N., Teregulov T. R.** Modeling a solar panel in MATLAB / Simulink, *Vestnik Ufimskogo gosudarstvennogo aviacionnogo tekhnicheskogo universiteta*, 2015,vol. 19, no. 4, pp. 70 (in Russian).

26. **Solar** Angle Calculator, available at: http://www.solarelectricityhandbook.com/solar-angle-calculator.html (Date of access: 5.03.2019).

УДК 004.932.2 DOI: 10.17587/it.26.602-607

В. А. Бобков, д-р техн. наук, гл. науч. сотр., e-mail: bobkov@iacp.dvo.ru,
В. П. Май, канд. техн. наук, вед. науч. сотр., e-mail: may@iacp.dvo.ru,
Институт автоматики и процессов управления ДВО РАН, г. Владивосток

Визуальная навигация автономного подводного робота с учетом самопересечений траектории[1]

Предложен метод визуальной навигации автономного подводного робота применительно к условиям локального маневрирования, ориентированный на повышение точности локализации робота за счет генерации виртуальной сети координатной привязки.

Ключевые слова: *визуальная одометрия, автономный подводный робот, навигация, loop closure*

Введение

Важной задачей при выполнении автономным подводным роботом (АПР) рабочей миссии в условиях априори неизвестной обстановки является задача его точной локализации, традиционно решаемая с помощью штатного навигационного оборудования. Более эффективному решению этой задачи способствует развитие подхода, основанного на визуальной одометрии. Такой подход особенно целесообразен в условиях локального маневрирования АПР, когда необходимы высокоточные перемещения АПР в ограниченной области. Однако известно, что для метода визуальной одометрии характерно накопление со временем ошибки вычисления траектории. Для уменьшения этой ошибки применяются различные методики: повторная инициализация расчета с привязкой к новой внешней системе координат (СК), метод выравнивания (bundle adjustment) [1] применительно ко всей траектории или локальное выравнивание [2], метод межкадровой привязки [3], интегрирование в вычислительную схему других сенсорных измерений [2] и др. В случае движения робота по самопересекающейся траектории возможно уточнение параметров траектории за счет использования преимуществ повторного посещения аппаратом/роботом одних и тех же мест (loop closure). В известных работах в этом направлении, например в работах [4—6], акцент делается на задаче опознавания мест. При этом одним из наиболее востребованных является так называемый "метод корзины слов" (bag-of-words method) [7]. Примером другого подхода, основанного на использовании предварительно подготовленных 3D-карт обстановки, является работа [8].

В настоящей статье предлагается развитие ранее разработанного метода визуальной навигации АПР, направленное на повышение точности локализации за счет учета самопересечений траектории при продолжительном маневрировании АПР в ограниченной области. Отличительной особенностью метода является построение на начальном этапе движения АПР виртуальной системы координатной привязки, состоящей из взаимосвязанных опорных систем координат.

Базовый метод визуальной навигации

Разработанный ранее базовый метод визуальной навигации [9, 2] (без учета самопересечений траектории) в соответствии с классической схемой реализации визуальной одометрии [10] содержит следующие шаги:

1. Выделение общего множества особенностей на четверке изображений двух стереопар. Сопоставление точечных особенностей выполняется на четырех изображениях — двух стереопарах (1-2 и 3-4), соответствующих двум последовательным позициям АПР на траектории. Для сопоставления особенностей используется детектор SURF (библиотека OpenCV) или трекер KLT. Для каждой пары изображений выполняется сопоставление слева направо и справа налево (cross-checking). Для исключения ложных сопоставлений применяется эпиполярный фильтр для пары 3-4 и для пары 1-2. Полученное в итоге множество особенностей,

[1] Работа выполнена при частичной финансовой поддержке РФФИ (проект № 18-07-00165), Программы "Дальний Восток" (проект 18-5-014) и Программы Президиума РАН "Фундаментальные проблемы решения сложных практических задач с помощью суперкомпьютеров".

сопоставленных для всех четырех изображений, позволяет далее построить два сопоставленных облака 3D-точек, отвечающих двум стереопарам (двум позициям АПР).

2. Генерация и фильтрация двух облаков 3D-точек.

3. Вычисление с помощью алгоритма ICP (Iterative Closest Point) локального геометрического преобразования H, связывающего локальные системы координат (СК) двух соседних позиций. Поиск этого преобразования основывается на имеющемся взаимно однозначном сопоставлении двух облаков (множеств) 3D-точек, наблюдаемых стереокамерой, соответственно, в позициях 1 и 2. Первое облако $C^1(x, y, z)$ задано в СК первой стереопары, второе облако $C^2(x, y, z)$ — в СК второй стереопары. Задача нахождения матрицы H формулируется как оптимизационная задача и решается с применением процедуры из библиотеки общего пользования MATLAB. В качестве параметров оптимизации используются три координаты вектора переноса и четыре координаты кватерниона, определяющего вращение в H. Ограничение задается условием равенства единице нормы кватерниона. Целевая функция — $F = \sum \left\| c_k^2 - c_k^1 H \right\|$, суммирование ведется по индексу k, где k — номер точки в облаке. Здесь $\{c_k^1\}$ — множество точек в первом облаке и $\{c_k^2\}$ — множество точек во втором облаке. Тогда с учетом вычисленной матрицы H новое положение АПР вычисляется через предыдущее как $p_2 = p_1 H$.

4. Вычисление параметров 6DOF (Six Degrees of Freedom) текущей позиции АПР в мировой системе координат (МСК) путем объединения последовательности локальных преобразований предшествующих позиций.

Метод навигации в условиях локального маневрирования

Предлагаемый метод визуальной навигации применительно к ситуации локального маневрирования АПР (когда движение осуществляется в ограниченной области подводной среды с возможными самопересечениями траектории) основывается на описанном выше и реализованном авторами базовом методе визуальной одометрии. Цель предлагаемого метода — воспрепятствовать накоплению ошибки навигации АПР при длительном локальном маневрировании за счет обработки ситуаций повторного посещения аппаратом одних и тех же мест и, тем самым, повысить точность навигации в целом. Входной информацией служит видеопоток, фиксируемый при движении АПР стереокамерой (направленной вниз). В работе метода выделяются два этапа, которые соответствуют двум этапам движения АПР по траектории. На первом этапе формируется сеть опорных систем координат (ОСК), которая при последующем движении АПР, т. е. на втором этапе, используется для повышения точности вычисления траектории АПР за счет возможных привязок к ОСК. В качестве ОСК рассматриваются локальные системы координат АПР/камеры в конкретных позициях траектории. Все ОСК связаны с мировой системой координат (МСК) через цепочку матриц геометрических преобразований, которые: а) порождаются непосредственно при работе базового метода (число матриц равно числу шагов) или б) берутся из уже существующей ОСК в случае успешной к ней координатной привязки (тогда число матриц равно числу матриц в ОСК + 1). Параметры траектории вычисляются в МСК, которая фиксируется в начальный момент времени. Поскольку накопление ошибки напрямую зависит от длины цепочки преобразований, то степень "точности" конкретной ОСК можно характеризовать степенью ее "близости" к МСК, т. е. можно присвоить каждой ОСК коэффициент степени близости k, равный длине цепочки преобразований, ведущей к МСК (если связь напрямую, то $k = 1$). Чем короче цепочка преобразований, тем меньшую ошибку при вычислении 6DOF она порождает. Этот коэффициент можно использовать для оптимального выбора ОСК на этапе навигации АПР в случае, когда АПР "видит" более одной ОСК.

Формирование ОСК

1. В начальный момент фиксируется МСК и стартует базовый метод визуальной навигации, с помощью которого на каждом шаге (шаг равен заданному числу кадров) выполняется вычисление параметров (6DOF) позиции траектории. Продолжительность траектории на этапе формирования ОСК предварительно задается отметкой времени или длиной пройденного аппаратом пути. Для более высокой эффективности использования сети ОСК желательно форму траектории этого участка выбирать таким образом, чтобы траектория по возможности равномерно и плотно покрывала (в плоскости дна) область маневрирования

АПР. Понятие плотности в данном случае подразумевает, что значительная площадь рассматриваемой области покрывается общей зоной видимости всех ОСК (камер в соответствующих позициях траектории).

2. В позициях траектории, определяемых заданным интервалом (числом кадров), осуществляется фиксация очередной новой $ОСК_{new}$, которая заносится в список ОСК, образующих сеть координатной привязки. С каждой ОСК связывается порция информации, необходимая для работы алгоритмов. В нее входит: стереопара снимков, полученная в данной позиции, абсолютные координаты АПР в данной позиции, накопленная в результате работы базового метода последовательность локальных матриц геометрического преобразования, геометрия видимого камерой участка дна. Если ближайшая (к текущей позиции) из уже существующих $ОСК_{old}$ имеет общую зону видимости с текущей позицией, то выполняется редактирование информации, относящейся к $ОСК_{new}$. А именно, ее последовательность локальных матриц заменяется на соответствующую последовательность матриц (более короткую), принадлежащую $ОСК_{old}$. Такая операция повышает точность локализации $ОСК_{new}$. Формирование ОСК выполняется также и на втором этапе, но только при условии их привязки к уже существующим ОСК. Следует заметить также, что эффективность виртуальной сети координатной привязки зависит от геометрии траектории АПР, поскольку все ОСК находятся на траектории.

Привязка к ОСК

Стандартная схема вычисления параметров каждой позиции траектории АПР в МСК предполагает согласно базовому методу перемножение всей цепочки матриц локальных преобразований, полученных на предыдущих шагах, что, как известно, приводит к накоплению существенной ошибки в случае длинной цепочки преобразований. Использование системы высокоточных ОСК позволяет потенциально исключать длинные цепочки преобразований на отдельных шагах, тем самым повышая точность навигации в целом. Суть работы алгоритма на этом этапе заключается в привязке текущей позиции к ОСК, когда это возможно (рис. 1), с заменой накопленной цепочки локальных преобразований на более короткую, принадлежащую ОСК, с дополнительным преобразованием H, связывающим текущую позицию с ОСК. Расчет перекрытия зон видимости, который лежит в основе проверки возможности привязки к ОСК, выполняется на основе известных данных о параметрах камеры и вычисленных параметрах траектории. Схема работы алгоритма представлена на рис. 2.

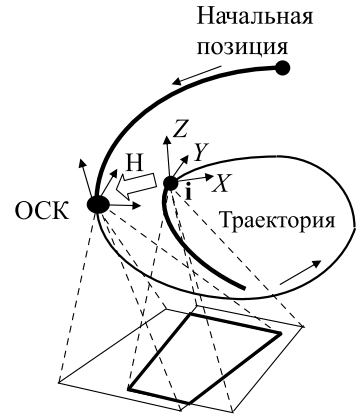


Рис. 1. Привязка АПР в текущей позиции *i* к ОСК: вычисление матрицы преобразования *H*; замена длинной цепочки преобразований (от ОСК до позиции *i* (участок траектории обозначен тонкой линией) на короткую (от начальной позиции до ОСК (участок траектории обозначен жирной линией) + преобразование *H*)

Рис. 2. Работа алгоритма при вычислении позиции АПР на текущем шаге

Режим мультисессии

В случаях, когда робот выполняет работу в одном и том же месте в разное время многократно, появляется возможность использовать для навигации виртуальную сеть координатной привязки, полученную в предыдущей

сессии. Это может дать дополнительное повышение точности навигации. Для этого необходимо обеспечить привязку новой траектории к прежней траектории. В нашем методе предполагается, что такую точную привязку можно осуществить базовым методом визуальной одометрии при условии, что начальная точка новой траектории будет находиться в близости к начальной точке прежней траектории. В данном случае понятие близости подразумевает, что зоны видимости двух камер должны перекрываться. Предварительные расчеты показали, что АПР, пользуясь штатными средствами навигации, может выйти в заданную точку с точностью, которая удовлетворяет указанному выше условию близости.

Таким образом, в повторных сеансах виртуальная сеть опорных систем координат становится гуще, что способствует дополнительному (по отношению к однопроходному режиму, когда виртуальная сеть привязки формируется и используется на протяжении одной траектории) повышению точности навигации АПР.

Обсуждение результатов

Эффективность рассмотренного метода в данной работе оценивали на модельных сценах. Помимо этого анализировали и аспекты практической реализации предлагаемых программно-алгоритмических средств, которые потенциально могут ограничивать применимость предложенной технологии, а именно:

1. *Достаточность вычислительных ресурсов АПР.* В некоторых современных автономных необитаемых подводных аппаратах (АНПА) в качестве вычислителей используются Atom 1 или 4 ядра, 1.6 ГГЦ, 1 ГБ ОЗУ. Внешняя память — твердотельные диски с объемом в среднем 500 Гбайт. На борту могут быть установлены 2...4 таких вычислителя, объединенных сетью Ethernet 100/1000. Для обработки изображений возможно использование Jetson Nvidia. Таких вычислительных ресурсов достаточно для работы предлагаемых программно-алгоритмических средств.

2. Негативно влияет на точность расчета траектории визуальным методом вертикальное маневрирование АНПА (быстрое изменение параметров 6DOF приводит к ухудшению сопоставления особенностей на изображениях и, как следствие, к ошибкам в расчете локальных матриц геометрического преобразования). Чтобы уменьшить такое влияние, можно считать, что координатная привязка осуществляется на малой скорости движения АНПА, т.е. в режиме зависания. Для реализации этого режима у АНПА имеются вертикальные и горизонтальные подруливающие устройства. Стабилизация осуществляется по 5 степеням свободы (курс, дифферент, продольные и поперечные перемещения). Перемещения в вертикальной плоскости фиксируются по датчику глубины (датчику давления) или высоты (эхолоту). Точность стабилизации 4...5 см.

Вместе с тем, обработку визуальной информации можно интегрировать с измерениями штатной бортовой навигационной системы, что позволяет повысить точность навигации визуального метода [2].

3. Поскольку для надежной привязки к пунктам виртуальной сети координатной привязки требуется достаточное число особенностей на обрабатываемых изображениях, желательно планировать траекторию так, чтобы она проходила над участками, где это условие выполняется. Это возможно, поскольку навигационная система АНПА позволяет осуществлять поисковые движения с точностью в несколько метров.

4. Безусловно, на эффективность применения метода визуальной навигации влияет прозрачность воды. Возможные пути снижения влияния этого фактора: уменьшение высоты движения АНПА до 1...1,5 м; метод визуальной навигации "включается" только при наличии порогового числа сопоставляемых особенностей на изображениях, в противном случае работает штатная навигационная система АНПА.

Тестирование разработанных прототипных программно-алгоритмических средств метода проводили на модельных сценах, получаемых с помощью моделирующего комплекса [11]. На рис. 3 в качестве примера показана одна из

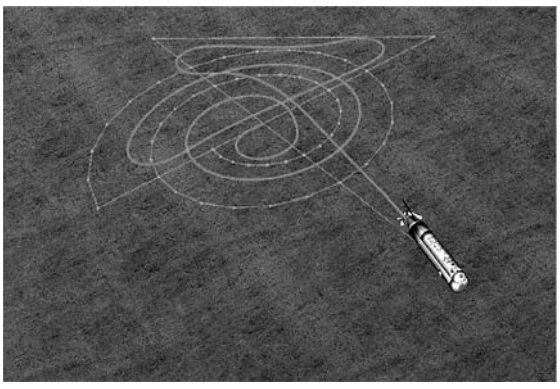


Рис. 3. Сцена для тестирования метода: траектория длиной 3180 кадров; частота съемки = 10 кадров/с; высота траектории над дном 0,6...3,3 м; шаг между расчетными позициями на траектории — 8 кадров

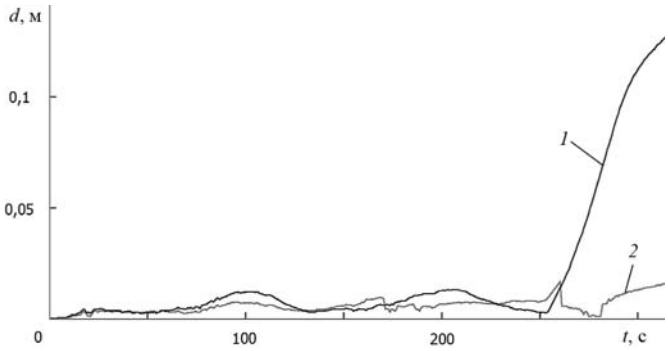


Рис. 4. Абсолютная ошибка расчета траектории АПР:
а — базовым визуальным методом (*1*); *б* — визуальным методом с использованием виртуальной сети координатной привязки (*2*). По оси *x* — время движения по траектории

виртуальных сцен. Результаты представлены на рис. 4. Как видно из графика, регулярная привязка к виртуальной сети ОСК предотвращает накопление погрешности при расчете траектории АПР. В процессе движения АПР на первом участке траектории было сгенерировано 33 ОСК, а при движении на втором участке было 37 обращений к ОСК. За счет привязки к ОСК накопление ошибки локализации АПР свелось к минимуму — 1,6 см вместо 13 см при обычной работе визуального метода.

В дальнейшем планируется развитие подхода как в части совершенствования алгоритмической базы, так и в части тестирования метода на реальных данных.

Заключение

Тестирование на модельных сценах предложенного метода визуальной навигации АПР с использованием виртуальной сети координатной привязки показало принципиальную возможность применения предложенной схемы вычисления параметров траектории АПР при длительных перемещениях робота в условиях локального маневрирования. В ближайшей перспективе планируется развитие метода в контексте решения задачи инспекции объектов подводной промышленной инфраструктуры.

Список литературы

1. **Triggs B., McLauchlan P., Hartley R., Fitzgibbon A.** Bundle Adjustment — A Modern Synthesis // ICCV '99: Proceedings of the International Workshop on Vision Algorithms. Springer-Verlag. 1999. P. 298—372.

2. **Bobkov V. A., Ron'shin Y. I., Kudryashov A. P., Mashentsev V. Y.** 3D SLAM from Stereoimages // Programming and Computer Software. 2014. Vol. 40, N. 4. *P.* 159—165.

3. **Бобков В. А., Борисов Ю. С. Навигация подводного аппарата на малых дистанциях по оптической информации** // Мехатроника, автоматизация, управление. 2010. № 2. С. 75—78.

4. **Olson E.** Recognizing places using spectrally clustered local matches // Robotics and Autonomous Systems. 2009. Vol. 57, N. 12. P. 1157—1172.

5. **Cummins M., Newman P.** FAB-MAP: Probabilistic Localization and Mapping in the Space of Appearance // The International Journal of Robotics Research. 2008. Vol. 27, N. 6. P. 647—665.

6. **Cadena C., Gálvez-López D., Tardós J., Neira J.** Robust place recognition with stereo sequences // IEEE Transaction on Robotics. 2012. Vol. 28, N. 4. P. 871—885.

7. **Sivic J., Zisserman A.** Video google: A text retrieval approach to object matching in videos // In Proceedings of the International Conference on Computer Vision. 2003. Vol. 2. P. 1470—1477.

8. **Pinto M., Moreira A. P., Matos A., Sobreira H., Santos F.** Fast 3D Map Matching Localisation Algorithm // International Conference on Computer and Automation Engineering (ICCAE 2013).

9. **Бобков В. А., Роньшин Ю. И., Машенцев В. Ю., Кудряшов А. П.** Навигация автономного подводного аппарата по видеопотоку // Информационные технологии. 2013. № 3. С. 36—41.

10. **Бобков В. А., Май В. П.** О повышении эффективности решения 3D SLAM задачи по стереоизображениям // Информатика и системы управления. 2018. № 2. С. 14—23.

11. **Melman S., Bobkov V., Inzartsev A., Pavin A.** Distributed Simulation Framework for Investigation of Autonomous Underwater Vehicles' Real-Time Behavior // Proceedings of the OCEANS'15 MTS/IEEE Washington DC, October 19—22, 2015.

V. A. Bobkov, Dr. Tech. Sc., Chief Scientific Researcher**,** e-mail: bobkov@iacp.dvo.ru,
V. P. May, Cand. Tech. Sc., Leading Scientific Researcher, e-mail: may@iacp.dvo.ru,
The Institute of Automation and Control Processes, Vladivostok, 690041, Russian Federation

Visual Navigation of Autonomous Underwater Robot with Loop Closing

The method of visual navigation of an autonomous underwater robot (AUV) for conditions of local maneuvering is described in article. The method aims to increase the accuracy of robot localization through generation and the using of virtual network for coordinate binding. The algorithms that implement the proposed method are based on the use of visual odometry and the proposed algorithms for re-visited places. The virtual coordinate binding network consists of the binding points generated at

the initial stage of the AUV movement. A binding point refers to a piece of data associated with a particular AUV position. This piece of data includes the position coordinates, a stereo pair of images visible by the camera from a given position, and a sequence of local geometrical transformation matrices accumulated by a given moment in time. It is assumed that the binding points have a high accuracy of coordinates since the method of visual navigation does not accumulate a large error with short movements of the AUV.At the stage of using the virtual network for coordinate binding a search and binding to the nearest binding point are performed. Improving the accuracy of AUV navigation is achieved by replacing a long sequence of local transformations associated with the current position with a shorter sequence associated with the binding point. Estimates of the effectiveness of the method on virtual scenes are obtained.

Keywords: *visual odometry, autonomous underwater vehicle, navigation, loop closing, virtual coordinate referencing network*

Acknowledgements: This work was supported by the Russian Foundation for Basic Research (project No. 18-07-00165), "Priority scientific research studies for comprehensive development of the Far-Eastern Division of the Russian Academy of Science" Program (project No. 18-5-014) and the Programs of the RAS Presidium "Fundamental problems of tackling complex practical tasks using supercomputers" (project "Supercomputer modeling of dangerous oceanic and other natural phenomena and control of technogenic objects").

DOI: 10.17587/it.26...

References

1. **Triggs B., McLauchlan P., Hartley R., Fitzgibbon A.** Bundle Adjustment — A Modern Synthesis, ICCV '99: Proceedings of the International Workshop on Vision Algorithms, Springer-Verlag, 1999, pp. 298—372.

2. **Bobkov V. A., Ron'shin Y. I., Kudryashov A. P., Mashentsev V. Y.** 3D SLAM from Stereoimages, *Programming and Computer Software*, 2014, vol. 40, no. 4, *pp.* 159—165.

3. **Bobkov V. A., Borisov Ju. S.** Underwater vehicle navigation on small distances from optical data, *Mehatronika, Avtomatizacija, Upravlenie*, 2010, no. 2, pp. 75—78 (in Russian).

4. **Olson E.** Recognizing places using spectrally clustered local matches, *Robotics and Autonomous Systems*, 2009, vol. 57, no. 12, pp. 1157—1172.

5. **Cummins M., Newman P.** FAB-MAP: Probabilistic Localization and Mapping in the Space of Appearance, *The International Journal of Robotics Research*, 2008, vol. 27, no. 6, pp. 647—665.

6. **Cadena C., Gálvez-López D., Tardós J., Neira J.** Robust place recognition with stereo sequences, *IEEE Transaction on Robotics*, 2012, vol. 28, no. 4, pp. 871—885.

7. **Sivic J., Zisserman A.** Video google: A text retrieval approach to object matching in videos, *In Proceedings of the International Conference on Computer Vision*, 2003, vol. 2, pp. 1470—1477.

8. **Pinto M., Moreira A. P., Matos A., Sobreira H., Santos F.** Fast 3D Map Matching LocalizationAlgorithm, *International Conference on Computer and Automation Engineering (ICCAE 2013),* Brussels, Belgium, 12—13 January, 2013.

9. **Bobkov V. A., Ron'shin Yu. I.,, Mashentsev V. Yu., Kudryashov A. P.** Navigation of an autonomous underwater vehicle by video stream, *Informacionnue Technologii*, 2013, no. 3, pp. 36—41 (in Russian).

10. **Bobkov V. A., May V. P.** On enhancement the solution efficiency of the 3D SLAM problem on stereo images, *Informatika i Sistemu Upravlenia*, 2018, no. 2, pp. 14—23 (in Russian).

11. **Melman S., Bobkov V., Inzartsev A., Pavin A.** Distributed Simulation Framework for Investigation of Autonomous Underwater Vehicles' Real-Time Behavior, *Proceedings of the OCEANS'15 MTS/IEEE Washington DC*, October 19—22, 2015.

Адрес редакции:

107076, Москва, Стромынский пер., 4

Телефон редакции журнала **(499) 269-5510**
E-mail: it@novtex.ru

Технический редактор *Е. В. Конова.*
Корректор *М. Ю. Безменова.*
Сдано в набор 10.08.2020. Подписано в печать 25.09.2020. Формат 60×88 1/8. Бумага офсетная.
Усл. печ. л. 8,86. Заказ IT1020. Цена договорная.

Журнал зарегистрирован в Министерстве Российской Федерации по делам печати, телерадиовещания и средств массовых коммуникаций.
Свидетельство о регистрации ПИ № 77-15565 от 02 июня 2003 г.

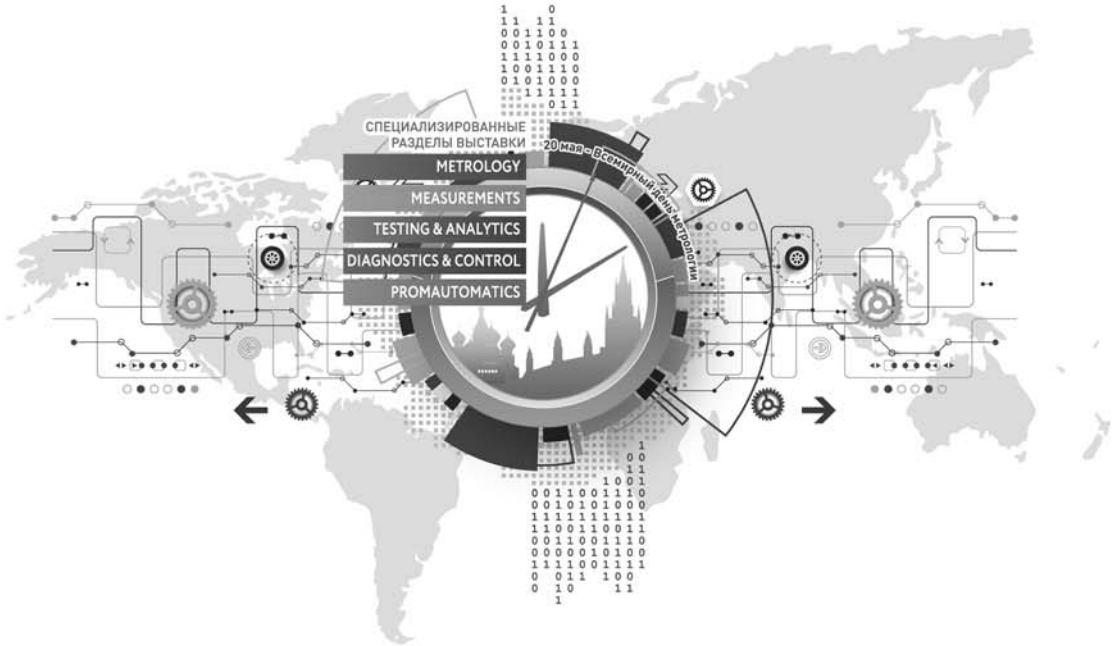
Оригинал-макет ООО "Адвансед солюшнз". Отпечатано в ООО "Адвансед солюшнз".
119071, г. Москва, Ленинский пр-т, д. 19, стр. 1. Сайт: **www.aov.ru**

16-й МОСКОВСКИЙ МЕЖДУНАРОДНЫЙ
ИННОВАЦИОННЫЙ ФОРУМ И ВЫСТАВКА

MetrolExpo'2020

ТОЧНЫЕ ИЗМЕРЕНИЯ –
ОСНОВА КАЧЕСТВА И БЕЗОПАСНОСТИ

1–3 декабря
Москва, ВДНХ, пав. 55



Новый гибридный формат выставки

офлайн + онлайн

Стирает границы
неограниченное количество участников со всего мира

Увеличивает охват
использование искуственного интеллекта для формирования рекомендаций и нетворкинга

Упрощает коммуникации
благодаря современным IT-технологиям

ДОСТУПНО В Google Play

Available on the App Store

Платформа представлена в связке классических веб-страниц и приложения для IOS и Android.

ОРГАНИЗАТОР:
Выставочная компания «ВЭСТСТРОЙ ЭКСПО»

Телефон/Факс: +7 (495) 937-40-23 (многоканальный)
E-mail: metrol@expoprom.ru

www.metrol.expoprom.ru

Рисунки к статье А. В. Савченко, И. С. Гречихина

«ДЕТЕКТИРОВАНИЕ СПЕЦИАЛИЗИРОВАННЫХ КАТЕГОРИЙ ОБЪЕКТОВ НА ФОТОГРАФИЯХ В МОБИЛЬНЫХ УСТРОЙСТВАХ НА ОСНОВЕ МНОГОЗАДАЧНОЙ НЕЙРОСЕТЕВОЙ МОДЕЛИ»

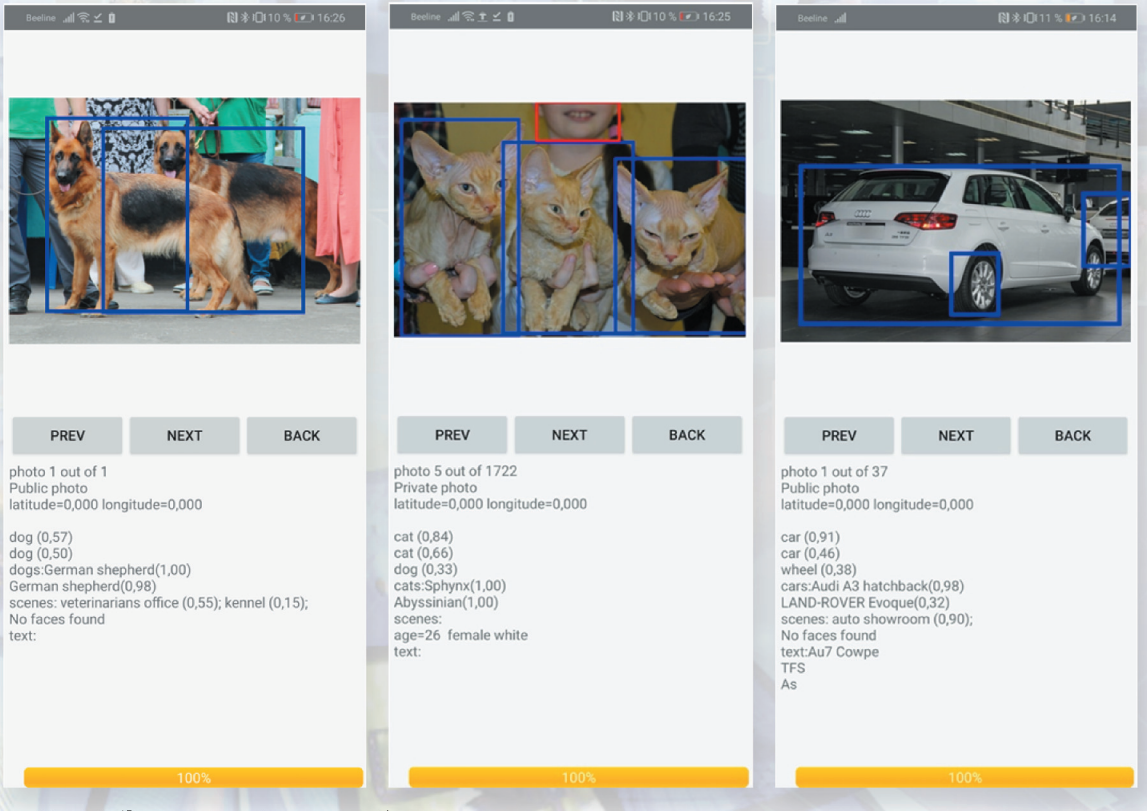


Рис. 4. Экранные формы мобильного приложения, реализующего предложенный подход

Рис. 5. Примеры изображений из наборов данных Stanford Dogs и Oxford-IIIT-Pet

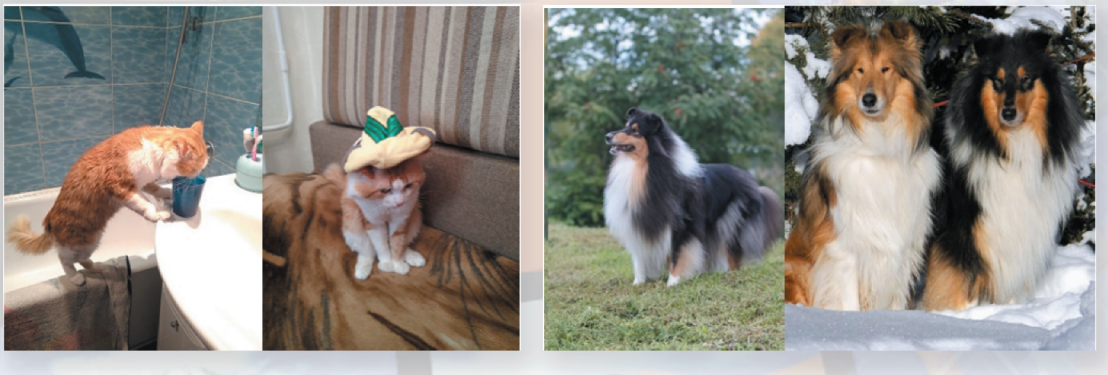


Рис. 6. Примеры изображений из собранного набора фотографий для тестирования качества кластеризации

Рисунки к статье С. В. Валевича, В. С. Осиповича, И. Крузе, Р. М. Асимова

«ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ МОНИТОРИНГА ТЕХНИЧЕСКОГО СОСТОЯНИЯ СОЛНЕЧНЫХ ЭЛЕКТРОСТАНЦИЙ»

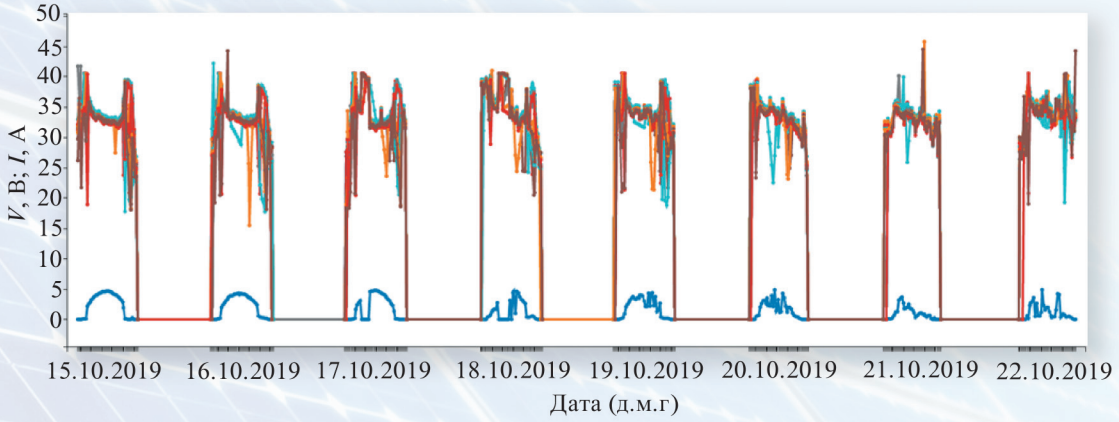


Рис. 3. Временные ряды напряжений и силы тока для группы последовательно подключенных панелей собранные в течение недели

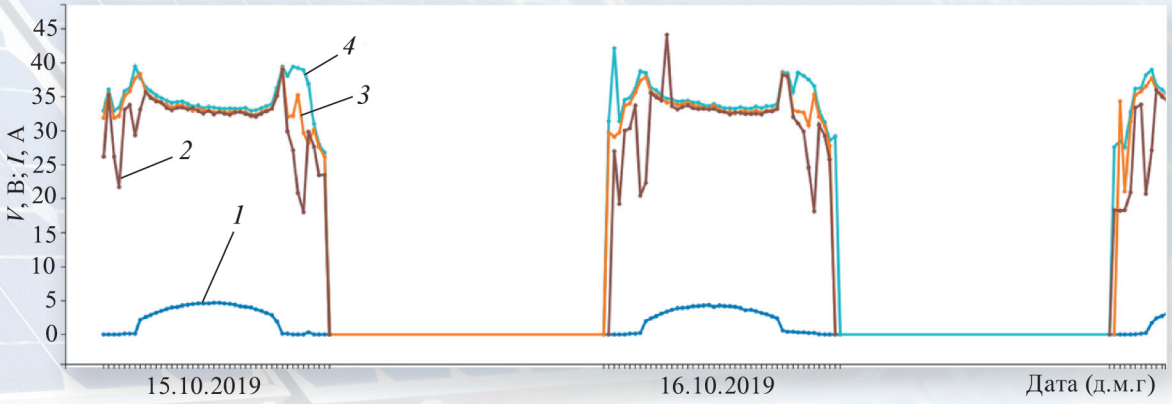


Рис. 5. Временные ряды силы тока (*1*) и напряжений солнечных панелей с номерами 16_11 (*2*), 16_18 (*3*), 16_17 (*4*)

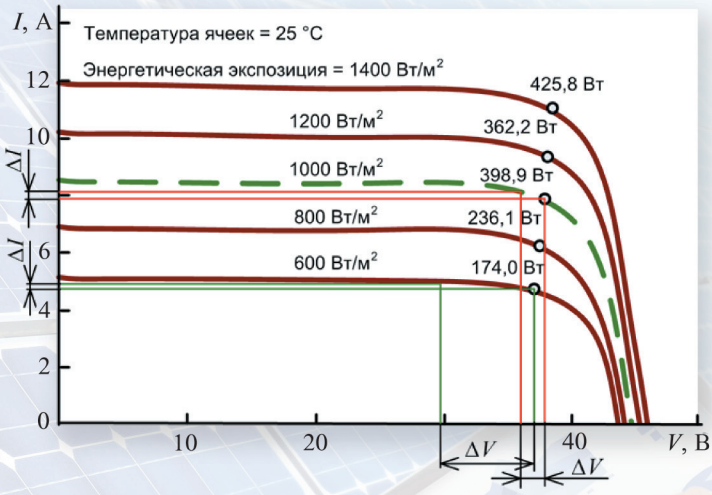


Рис. 6. Образец вольт-амперных характеристик солнечной панели при различных уровнях энергетической экспозиции

ISSN 1684-6400. Информационные технологии. 2020. Том 26. № 10. 553–608. Индекс: 72656 (Объединенный каталог «Пресса России»)