

УДК 004.056.53

О. А. ЛАПТЄВ, канд. техн. наук, ст. наук. співробітник;

Г. В. ШУКЛІН, канд. техн. наук;

В. А. САВЧЕНКО, доктор техн. наук, професор,

Державний університет телекомунікацій, Київ

МЕТОД ОЦІНЮВАННЯ ПАРАМЕТРІВ ІМПУЛЬСНОГО СИГНАЛУ НА ОСНОВІ КОРЕЛЯЦІЙНО-РЕГРЕСІЙНОГО АНАЛІЗУ

Розроблено математичну модель статистичного оцінювання параметрів імпульсного сигналу на основі регресійного аналізу, який набув широкого застосування для визначення параметрів, котрі ідентифікують засоби негласного отримання інформації (закладки), з метою оцінювання ймовірності відшукування сигналів від прихованих пристроїв, що відрізняються один від одного за технічними характеристиками та принципом дії. Регресійний аналіз дозволяє визначити емпіричну формулу, що описує залежність виявлення сигналу, джерелом якого є засоби негласного отримання інформації від параметрів, завдяки яким розпізнаються типи пристроїв. За допомогою кореляційно-регресійного методу розроблено методику оцінювання взаємозв'язку сигналів від засобів негласного отримання інформації, що дає можливість моделювати процеси пошуку засобів негласного знімання інформації (закладок) за наявності багатьох змінних.

Дану методику підтверджено за результатами застосування критеріїв Фехнера та Пірсона для здобутих експериментальних даних. Обидва критерії оцінювання дали практично однакові висновки стосовно вірогідності моделювання, що свідчить про адекватність моделі та повністю відповідає практичному аспекту пошуку закладок.

Ключові слова: регресійний аналіз; сигнал; матриця; коваріація.

Вступ

Математичні моделі об'єктів і процесів реальної дійсності ґрунтуються на використанні теоретико-аналітичного та експериментально-статистичного підходах.

Застосування теоретико-аналітичного підходу можливо за наявності необхідної інформації для синтезу формалізованого опису. З ускладненням розв'язуваних завдань, появи нових систем, що генерують інноваційні процеси, розробка методик і методологій на основі виключно теорій стає дуже складною, важкою і не достатньо точною. Реальним джерелом для побудови математичних моделей є статистичні дані, здобуті експериментально.

Теоретично — це математичний підхід, котрий здебільшого своєю реалізацією спрощує дійсність та, узагальнюючи, ігнорує особливості та індивідуальність. Експериментально-статистичний підхід будує свої моделі на індивідуальних, конкретних даних, які не відривають об'єкт за межі часу і місця. Вони використовують методологію статичного опису і формулюють через поодинокі прояви загальні закономірності та закони поведінки безлічі об'єктів, які в масових явищах стають імовірними.

Отже, статистичні дані, які є результатами експериментів, створюють актуальні початкові умови для вивчення методів і процесів, що дають можливість імовірно оцінювати вірогідність конкретних методів і процесів.

Короткий аналіз літературних джерел. Більшість відомих підходів до моделювання різняться тим, що залежно від використовуваного математичного апарата (теорія ймовірності, випадкові процеси, теорія графів, нечітких множини та ін.), розраховуються різні параметри під час визначення вхідної та вихідної інформації самої моделі. При цьому аналітичні моделі, що розглядаються з позиції теоретичної математики, не тотожні реальній дійсності через обмежену точність результатів [1; 3].

У [4] досліджено модель інформаційної безпеки на основі марковських випадкових процесів. Здобуто чисельні значення, проте розкрито лише питання щодо загрози уразливості.

У [5] приділено увагу нестійкості, тобто маємо чималу кількість варіацій здобутих розв'язків із поганою узгодженістю систем лінійних алгебраїчних рівнянь і неточно заданих значень ефектів та результатів спостережень.

Водночас у багатьох джерелах математичне моделювання розглядається як математична модель конкретних параметрів, що можуть бути як детерміновані, так і випадкові. Однак взаємозв'язок між вхідними параметрами у процесі моделювання не розглядається. Ці чинники взаємозв'язку і взаємозвпливу можуть істотно спотворити результати моделювання і поставити під сумнів адекватність моделі.

Таким чином, оцінювання параметрів сигналу засобу негласного отримання інформації (ЗНОІ), взаємний вплив одного характерного параметра засобу негласного знімання інформації на інший, залеж-

ність цих параметрів від кількості проведених практичних заходів є дуже важливими на сучасному етапі.

Постановка проблеми. У процесі пошуку ЗНОІ постає завдання з визначення основного переліку параметрів і властивостей засобів, які уможливають виявлення закладки з подальшим її локалюванням. Установлення мінімальної кількості параметрів ЗНОІ, які дають змогу їх ідентифікувати, безпосередньо залежить від кількості взаємозв'язаних параметрів. У разі вилучення на першому етапі взаємозв'язаних параметрів значно скорочується час пошуку ЗНОІ. Зважаючи на велику кількість взаємопов'язаних параметрів, кількісне їх оцінювання здійснюється за допомогою регресійного аналізу, тоді як якісне оцінювання можна отримати кореляційно-регресійним аналізом. Тому питання розробки методу оцінювання параметрів сигналу ЗНОІ на основі регресійного і кореляційно-регресійного аналізу сьогодні набуло особливої актуальності.

Основна частина

Метод регресійного аналізу — це найбільш довершений з існуючих методів аналізу статистичних даних, отриманих експериментальним шляхом. Він дає можливість ефективно проводити розрахунки з використанням сучасних інформаційних технологій і систем [2]. Застосування цього методу передбачає такі основні етапи:

- визначення ознак, що класифікують закладки;
- визначення параметрів, які є інформативними під час визначення ймовірності виявлення;
- вибір і обґрунтування характеру зв'язку виявлення закладки у разі зміни параметрів, що її характеризують;
- побудова системи нормальних рівнянь і розрахунок коефіцієнтів регресії.

Основним класифікатором ознаки закладки є параметричний ряд, елементи якого можуть створювати впорядковану множину, яка характеризує її тип залежно від застосування, умов і вимог експлуатації тощо. На етапі формування параметричних рядів можна скористатися методами автоматичної класифікації, які дають змогу із загальної множини закладок виокремлювати однорідні групи — класи еквівалентності.

Відбір параметрів, які ідентифікують закладки, здійснюється за такими основними вимогами:

- до складу вибраних уводять параметри, зафіксовані в стандартах радіозв'язку, а також такі, що характеризують ознаки радіозакладок;
- визначені параметри за своєю сукупністю повинні повною мірою характеризувати конструктивні, технологічні та експлуатаційні властивості закладок, що належать до ряду, і мати досить тісний кореляційний зв'язок із параметрами радіозакладок, які виявляються;
- самі визначені параметри не повинні бути взаємозалежні.

Для відбору техніко-економічних параметрів, які істотно впливають на виявлення закладок, обчислюється матриця коефіцієнтів парної кореляції. За значенням коефіцієнтів кореляції між параметрами можна робити висновок про тісноту зв'язку. Якщо визначений параметр істотно впливає на виявлення радіозакладок, то відповідний коефіцієнт кореляції близький до нуля. У загальному випадку система, з якою експериментують для отримання моделі, має вигляд, як це зображено на рис. 1, де X — керовані фактори, які можна змінювати та підтримувати на певних рівнях. На етапі пошуку ЗНОІ це можуть бути рівень перевищення порогу амплітуди, види модуляції, наявність другої і третьої гармонік і тип;

Z_1, Z_2, \dots, Z_k — це некеровані фактори, до яких належать режими роботи радіозакладок, тобто параметри, які контролюються ззовні чи не контрольовані системою;

G_1, G_2, \dots, G_k — не контрольовані установки та параметри радіомоніторингу або радіофіру;

Y — параметри виявлення як теоретичні, так і експериментальні.

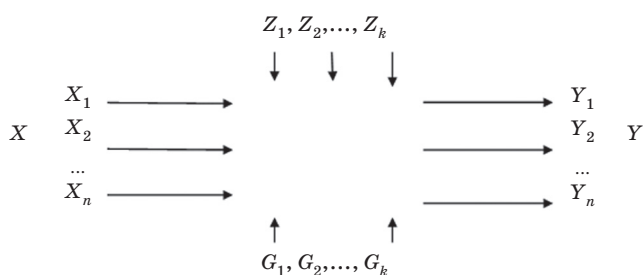


Рис. 1. Модель складної системи

Система — це математичне перетворення різних вхідних параметрів у вигляд, придатний для цифрового аналізу.

Щоб визначити емпіричну формулу, яка описує залежність виявлення сигналу, джерелом якого є ЗНОІ, від параметрів, за допомогою яких розпізнаються типи пристроїв, будемо використовувати модель лінійного виду

$$Y = B_0x_0 + B_1x_1 + \dots + B_kx_k + \varepsilon_l = \sum_{i=0}^k B_i x_i + \varepsilon_l, \quad (1)$$

де Y — параметр виявлення закладки; B_1, B_2, \dots, B_k — коефіцієнти регресії; x_1, x_2, \dots, x_k — керовані фактори, які задаються на вхід; k — загальна кількість факторів; ε_l — випадкова помилка для l -го досліджу.

Значення помилки обчислюватимемо за співвідношенням

$$\varepsilon_l = Y - M(Y), \quad (2)$$

де $M(Y)$ — математичне сподівання параметра виявлення Y .

Коефіцієнти регресії B_1, B_2, \dots, B_k розраховуватимемо за методом найменших квадратів, який полягає в мінімізації суми квадратів відхилень значень \hat{Y} , здобутих розрахунково, і значень Y , отриманих дослідним шляхом, тобто:

$$\sum_1^N [Y - \hat{Y}]^2 = \sum_1^N [Y - (b_0 x_0 + b_1 x_1 + \dots + b_k x_k)]^2 = Q = \min. \quad (3)$$

Тут значення $b_1 \approx B_1, b_2 \approx B_2, \dots, b_k \approx B_k$ дістаємо, беручи частинні похідні функції Q за b_1, b_2, \dots, b_k , прирівнюємо їх до нуля і розв'язуємо систему лінійних рівнянь:

$$\begin{cases} \frac{\partial Q}{\partial b_0} = 0; \\ \frac{\partial Q}{\partial b_1} = 0; \\ \dots \\ \frac{\partial Q}{\partial b_k} = 0. \end{cases} \quad (4)$$

Тоді рівняння (1) набирає вигляду

$$\hat{Y} = B_0 x_0 + B_1 x_1 + \dots + B_k x_k = \sum_{i=0}^k B_i x_i. \quad (5)$$

Це є регресійна модель пошуку засобів негласного отримання інформації.

Залежність (5) дає можливість дістати густину розподілу ймовірності поширення випадкових величин x та y . У нашому разі рівняння регресії, яке визначає залежність y від x буде лінійним. Тоді функція розподілу ймовірності двовимірної випадкової величини (x, y) буде мати вигляд

$$F_i(x, y) = \frac{x_i y_i}{x_{\max} y_{\max}}. \quad (6)$$

Отже, густина розподілу ймовірності визначатиметься за виразом

$$f_i(x, y) = \frac{\partial F_i(x, y)}{\partial x y} = \frac{1}{x_{\max} y_{\max}}. \quad (7)$$

Обчисливши ймовірності для кожної пари значень x_i та y_i , дістанемо мінімальні значення, якими в подальшому під час вибору параметрів визначення ЗНОІ можна знехтувати.

З метою скорочення обчислень розрахунки виконувалися за кореляційно-регресійним методом для визначення коефіцієнтів Фехнера та Пірсона, які характеризують зв'язок вхідних і вихідних параметрів.

Щоб оцінити тісноту зв'язку в теорії кореляції застосовується шкала англійського статистика Чеддока: слабка — від 0,1 до 0,3; помірна — від 0,3 до 0,5; помітна — від 0,5 до 0,7; висока — від 0,7 до 0,9; вельми висока (тісна) — від 0,9 до 1,0.

У контексті даної статті кореляція характеризує лінійний взаємозв'язок у варіаціях змінних. Вона може бути парною (корелюють дві змінні — частота сканування та діапазон сканування) або множинної (більш як дві змінні), прямої або зворотної — позитивної або негативної, коли змінні варіюють відповідно в однакових або різних напрямках.

У нашому разі частота сканування і амплітуда перевищення сигналу файла зразка є кількісними змінними і рівноцінні у своїх незалежних спостереженнях при їх загальній кількості. При цьому емпіричними заходами тісноти їх лінійного взаємозв'язку є коефіцієнт прямої кореляції знаків Г. Т. Фехнера і коефіцієнти парної, частинної та множинної кореляції К. Пірсона.

Коефіцієнт парної кореляції знаків Фехнера визначає узгодженість напрямів в індивідуальних відхиленнях змінних від своїх середніх значень. Він дорівнює відношенню різниці сум, що збігаються (Mat), до сум, що розбігаються (Mit), пар знаків у відхиленнях $\varepsilon_y = y - \bar{y}$ та $\varepsilon_x = x - \bar{x}$ до суми цих сум:

$$K_f = \frac{\sum_1^N Mat - \sum_1^N Mit}{\sum_1^N Mat + \sum_1^N Mit}. \quad (8)$$

Тут K_f — коефіцієнт Фехнера; Mat (*Matching* — збіжність); Mit (*Mitmatching* — розбіжність).

Коефіцієнт K_f , як зазначалося, змінюється від -1 до $+1$. Сумування в (8) виконувалося за спостереженнями. Якщо деяке одне відхилення нульове, то ним можна знехтувати. Якщо ж обидва відхилення нульові, то такий випадок вважається збіжним за знаками і його беруть до розрахунку. Статистичні дані обчислення коефіцієнта Фехнера наведено в табл. 1.

Таблиця 1

Номер перевірки N	Кількість сканувань x	Виявлення перевищення амплітуди y	Відхилення від середніх		Порівняння	
			$x = 41$	$y = 43$	Mat	Mit
1	55	43	14	0		
2	45	42	4	-1	0	1
3	35	39	-6	-4	1	0
4	25	40	-16	-3	1	0
5	35	42	-6	-1	1	0
6	45	47	4	4	1	0
7	50	53	9	10	1	0
8	40	38	-1	-5	1	0
9	35	45	-6	2	0	1
10	45	41	4	-2	0	1
Разом	410	430			6	3

Таблиця 2

Номер перевірки N	Кількість сканувань x	Виявлення перевищення амплітуди y	Розрахунки		
			x^2	y^2	xy
1	55	43	3025	1849	2365
2	45	42	2025	1764	1890
3	35	39	1225	1521	1365
4	25	40	625	1600	1000
5	35	42	1225	1764	1470
6	45	47	2025	2209	2115
7	50	53	2500	2809	2650
8	40	38	1600	1444	1520
9	35	45	1225	2025	1575
10	45	41	2025	1681	1845
Разом	410	430	$\Sigma x^2 = 17\ 500$	$\Sigma y^2 = 18\ 666$	$\Sigma xy = 7795$

Слід зазначити, що дані в табл. 1 заносили на основі реальних перевірок приміщень на наявність засобів негласного знімання інформації (радіозакладок).

Із розрахунків випливає, що коефіцієнт Фехнера $K_f = 0,33$, тобто згідно зі шкалою англійської статистика Чеддока залежність під час даних спостережень вийшла помірною (що повністю відповідає практичним результатам), кількість сканувань частотного діапазону явно не впливає на виявлення сигналу і перевищує за амплітудою файл зразка, тобто безпосередньо не позначається на виявленні ЗНОІ.

Якщо в процесі аналізу необхідно враховувати відхилення змінних, то слід застосовувати коефіцієнти парної, частинної та множинної лінійної кореляції Пірсона. Для їх розрахунку використовують різні методи. Ми послугуємося методом прямого рахунку. Тоді коефіцієнт парної кореляції Пірсона набере вигляду

$$r_{xy} = r_{yx} = \frac{N \sum xy - \sum x \sum y}{\sqrt{(N \sum x^2 - (\sum x)^2)(N \sum y^2 - (\sum y)^2)}}. \quad (9)$$

Для коректного порівняння беремо початкові дані з табл. 1 і виконуємо розрахунки. Результати попередніх розрахунків та початкових умов подано в табл. 2.

Для аналізу до здобутих розрахунків, наведених у табл. 2, додаємо значення відхилення і, скориставшись формулою (9), обчислюємо коефіцієнт парної кореляції Пірсона. Отже, маємо $r_{xy} = 0,4735$, що згідно зі шкалою Чеддока відповідає помірній залежності і повністю підтверджує результати обчислень, раніше проведених у даній статті для оцінювання залежності досліджуваних параметрів за допомогою коефіцієнта Фехнера.

Напрямок подальших досліджень є визначення чисельних оцінок багатовимірної регресійної моделі пошуку засобів негласного знімання інформації (радіозакладок) стосовно її статистичної надійності або, інакше кажучи, щодо адекватності моделі, яка описує процес.

З огляду на викладене можна дійти висновку, що розроблений метод оцінювання виявлення сигналу радіозакладки на основі кореляційно-регресійного аналізу підтверджується двома різними методами оцінювання, що свідчить про високу якість запропонованої моделі.

Висновки

◆ Розроблено математичну модель регресійного аналізу процесу пошуку пристроїв негласного отримання інформації (закладок) для багатовимірного процесу. Така модель дає можливість оцінити взаємовплив вхідних сигналів і вихідних факторів визначення засобів негласного отримання інформації.

◆ На основі реальних статистичних даних за допомогою цієї математичної моделі та кореляційно-регресійного аналізу процесу пошуку засобів негласного отримання інформації здійснено оцінювання впливу двох вхідних факторів на визначення пристрою негласного отримання інформації.

◆ З'ясовано, що з двох вибраних вхідних параметрів тільки один істотно впливає на процес виявлення пристрою негласного отримання інформації.

♦ Оцінку адекватності здобутих даних було перевірено за двома різними методами: критеріями Фехнера та Пірсона. Обидва методи оцінювання дали практично однакові результати, що підтверджує адекватність моделі та повністю відповідає практиці пошуку засобів негласного отримання інформації.

Список використаної літератури

1. **Регрессионный анализ** [Электронный ресурс]. URL: <http://www.machinelearning.ru/> (01.05.2019).
2. **Методы** изучения взаимосвязи социально-экономических явлений с помощью корреляционно-регрессивного анализа [Электронный ресурс]. URL: www.Grandars.ru » Статистика » Общая теория статистики » (01.07.2019).
3. **Лаптев О. А., Степаненко В. І., Тихонов Ю. О.** Формальні математичні моделі для забезпечення безпеки інформації // Сучасний захист інформації. 2019. №1(37). С. 59–64.
4. **Лаптев О. А.** Модель інформаційної безпеки на основі марковських випадкових процесів // Зв'язок. 2018. №6 (136). С. 43–48.
5. **The method** of searching for digital means of illegal reception of information in information systems in the working range of Wi-Fi / A. A. Laptev, O. V. Barabash, V. V. Savchenko [et al.] // International Journal of Advanced Research in Science, Engineering and Technology. India. 2019. Vol. 6. Issue 7. P. 10101–10105.
6. **The method** of searching for digital means of illegal obtaining of information on the basis cluster analysis / A. A. Laptev, V. A. Savchenko, O. V. Barabash [et al.] // Magyar Tudományos Journal. Budapest, Hungary. 2019. № 31. P. 33–37.
7. **Радченко С. Г.** Методология регрессионного анализа: монография. Київ: Корнійчук, 2011. 375 с.
8. **Методы** корреляционного и регрессионного анализа [Электронный ресурс]. URL: <https://studme.org/33796/informatika/metody/> (07.07.2019).
9. **Методы** сравнительной комплексной оценки хозяйственно-финансовой деятельности организации [Электронный ресурс]. URL: <https://lesnaya.nethouse.ru/articles/33525/> (27.07.2019).

Рецензент: доктор техн. наук, професор **О. В. Барабаш**, Державний університет телекомунікацій, Київ.

А. А. Лаптев, Г. В. Шуклин, В. А. Савченко

МЕТОД ОЦЕНКИ ПАРАМЕТРОВ ИМПУЛЬСНОГО СИГНАЛА НА ОСНОВЕ КОРРЕЛЯЦИОННО-РЕГРЕССИВНОГО АНАЛИЗА

Разработана математическая модель статистической оценки параметров импульсного сигнала на основе регрессивного анализа, получившего широкое применение для определения параметров, которые идентифицируют средства негласного получения информации (закладки), с целью оценки вероятности обнаружения сигналов от скрытых устройств, которые отличаются друг от друга техническими характеристиками и принципом действия. Регрессивный анализ позволяет определить эмпирическую формулу, описывающую зависимость обнаружения сигнала, источником которого являются средства негласного получения информации от параметров, благодаря которым распознаются типы устройств. С помощью корреляционно-регрессивного метода разработана методика оценки взаимовлияния сигналов от средств негласного получения информации, что дает возможность моделировать процессы поиска средств негласного снятия информации (закладок) при наличии многих переменных.

Данная методика подтверждена результатами применения критериев Фехнера и Пирсона для полученных экспериментальных данных. Оба критерия оценки дали практически одинаковые выводы относительно достоверности моделирования, что свидетельствует об адекватности модели и полностью соответствует практическому аспекту поиска закладок.

Ключевые слова: регрессивный анализ; сигнал; матрица; ковариация.

O. Laptev, G. Shuklin, V. Savchenko

METHOD OF ASSESSMENT OF PULSE SIGNAL PARAMETERS BASED ON CORRELATION REGRESSION ANALYSIS

The mathematical model of statistical estimation of impulse signal parameters on the basis of regression analysis is developed. characteristics and principle of action.

Theoretical — mathematical approach in most cases of its realization simplifies reality, generalizes, ignores features and individuality. The experimental-statistical approach builds its models on individual, concrete data that does not detach the object from the time and place frames. They use the methodology of static description and formulate, through single manifestations, the general patterns and laws of behavior of many objects that become probable in mass phenomena. The real method of obtaining mathematical models is experimentally statistical approaches. That is, the main sources of raw data is a really accumulated static database of a particular type of process or method under study.

Regression analysis allows us to determine an empirical formula that describes the dependence of detecting a signal, the source of which is the means of silently obtaining information from the parameters by which the types of devices are recognized.

Using the correlation-regression method, a technique for estimating the mutual influence of signals from the means of silent retrieval was developed, which makes it possible to simulate the processes of finding the means of silent retrieval of information (bookmarks) in the presence of many variables.

This technique is confirmed by the results of applying the Fechner and Pearson criteria to the experimental data obtained. Both evaluation criteria gave virtually the same conclusions as to the reliability of the simulation, which confirms the adequacy of the model and is fully consistent with the practical aspect of bookmarking.

Keywords: regression analysis; signal; matrix; covariance.