

ЕКСПЕРИМЕНТАЛЬНО-СТАТИСТИЧНИЙ МЕТОД ОБЧИСЛЕННЯ КОРЕЛЯЦІЙНОЇ ВЗАЄМОЗАЛЕЖНОСТІ ПАРАМЕТРІВ РОЗПІЗНАВАННЯ ЗАСОБІВ НЕГЛАСНОГО ОТРИМАННЯ ІНФОРМАЦІЇ

Розглядається експериментально-статистичний підхід створення методів обчислення параметрів взаємозалежності параметрів розпізнавання засобів негласного отримання інформації на індивідуальних, конкретних даних, які не відривають об'єкт від рамок часу і місця. На основі кореляційно-регресійного аналізу, запропоновано метод оцінки взаємовпливу сигналів засобів негласного отримання інформації, для лінійної залежності параметрів розпізнавання засобів негласного отримання інформації двома різними методами, визначенням коефіцієнту Фехнера та коефіцієнту Пірсона. Які показали практично однакові результати, що свідчить про достовірності результатів і правильності запропонованої методики оцінки численних значень параметрів взаємовпливу. За допомогою запропонованого алгоритму визначення коефіцієнта рангової кореляції Спірмена, визначили параметри розпізнавання засобів негласного отримання інформації з нелінійною залежністю, отримані результати підтверджують відсутність взаємозалежності запропонованих параметрів розпізнавання засобів негласного отримання інформації. Обчисленим t-Критерієм Стьюдента, для коефіцієнта Спірмана, підтвердили адекватність запропонованого методу оцінки взаємовпливу параметрів розпізнавання засобів негласного отримання інформації.

Приведені у статті методи та розрахунки дозволяють вибрати самі основні параметри розпізнавання, що дозволить на першому етапі скоротити час розпізнавання, що дозволить збільшити ймовірність знаходження засобів негласного отримання інформації.

Ключові слова: кореляція, коефіцієнт, взаємозалежність параметрів, сигнал, ранг.

Вступ

Математичні моделі об'єктів і процесів реальної дійсності отримують з використанням теоретично-аналітичного та експериментально-статистичного підходів.

Використання теоретично-аналітичного підходу можливо, якщо маємо всю необхідну інформацію для синтезу формалізованого опису. З ускладненням завдань, що вирішуються, появи нових систем, процесів теоретичного рішення стає дуже складним, важким і до кінцевого рахунку недостатньо точним. Реальним методом отримання математичних моделей стає експериментально статистичні підходи. У якості даних можуть використовуватися результати проведених заходів, експертні оцінки. Тобто основними джерелами вихідних даних є реально накопичена статична база з того чи іншого виду досліджуваного процесу чи методу.

Теоретично — математичний підхід в більшості випадком своєї реалізації спрощує дійсність, узагальнюючи, ігнорує особливості та індивідуальність. Експериментально-статистичний підхід будує свої моделі на індивідуальних, конкретних даних, які не відривають об'єкт від рамок 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 ймовірними. Але для побудови моделі потрібно робить те чи інші припущення та виключати з моделі параметри які є взаємозамінні, з цією метою визначаються незалежні параметри моделювання.

Виходячи з вищевикладеного процес визначення сили взаємодії параметрів моделі одних на інші, їх взаємозалежність є дуже актуальним.

Короткий аналіз літературних даних

Більшість відомих підходів до моделювання, відрізняються тим, які параметри при моделюванні ними використовують в якості вхідної інформації та які характеристики модельованої системи розраховуються та надходять на вихід моделі (будують моделі з Використання Теорії ймовірностей, випадкових процесів, мереж Петрі, Теорії автоматів, Теорії графів, нечітких множини, Теорії катастроф, ентропійного підходу та ін.),

При цьому аналітичні моделі, що розглядаються з позиції теоретичної математики не тотожні реальної дійсності, зважаючи на обмежену точність результатів [1, 2].

У [2] розглядається модель інформаційної безпеки на основі Марковських випадкових процесів. Отримані чисельні значення, однак вони розглядають питання загрози уразливості. Питання загроз уразливості не торкається питання взаємозалежності основних параметрів моделі, що можливо призводить до ускладнення моделювання процесу.

У [3] звертається увага на нестійкість і отже, великі варіації отриманих рішень при поганій обумовленості систем лінійних алгебраїчних рівнянь і неточно заданих значень ефектів і результатів спостережень. Це пов'язане з питанням не врахування взаємозалежності основних параметрів

Разом з тим у всіх зазначених джерелах математичне моделювання розглядається як математична модель конкретних параметрів (деякі параметри мають імовірнісний характер) Питання взаємозв'язку вхідних параметрів при моделюванні процесів глибину їх взаємозв'язку моделі не розглядають. Ці чинники взаємозв'язку і взаємовпливу можуть істотно спотворити результати моделювання і поставити під сумнів адекватність моделі.

Виходячи з чого оцінка параметрів сигналу засобів негласного отримання інформації (ЗНОІ), взаємний вплив одного характерного параметра на інший, залежність цих параметрів від числа проведених практичних заходів є дуже важливою на сучасному етапі.

Постановка проблеми

В процесі пошуку ЗНОІ виникає задача визначення основного переліку параметрів і властивостей, за якими ЗНОІ виявляється і надалі локалізується. Виявлення мінімальної кількості параметрів ЗНОІ які дозволяють їх ідентифікувати безпосередньо залежить від кількості взаємопов'язаних параметрів. при виключенні на першому етапі взаємопов'язаних параметрів значно скорочується час пошуку ЗНОІ. Кількісну оцінку взаємопов'язаних параметрів дозволяє зробити кореляційно регресивний аналіз. Тому питання розробки методу оцінки параметрів сигналу ЗНОІ на основі кореляційно-регресійного аналізу є дуже актуальним.

Мета

Розробити кількісний метод оцінки параметрів сигналів ЗНОІ на основі кореляційно-регресійного аналізу, з метою відбору необхідної кількості головних параметрів не взаємопов'язаних між собою які дозволять розпізнати ЗНОІ з дуже великою ймовірністю.

Виклад основного матеріалу.

Для моделі визначення сигналу ЗНОІ доповнимо регресійний аналіз ще й кореляційним аналізом. Даний метод містить дві складові частини — кореляційний аналіз та регресійний аналіз. Кореляційний аналіз — це кількісний метод визначення тісноти та напрямку взаємозв'язку між вибірковими змінними величинами. Регресійний аналіз та регресійна модель ґрунтується на цьому методі детально розглянемо вище і визначає кількісний метод визначення виду математичної функції в причинно-наслідкової залежності між змінними величинами.

Для оцінки сили зв'язку в теорії кореляції застосовується шкала англійської 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 (сильна) - від 0,9 до 1,0.

При нашому аналізі кореляція характеризує лінійний взаємозв'язок у варіаціях змінних.

Вона може бути парною (дві корелюють змінна-частота сканування та діапазон сканування) або множинної (більш як двоє змінних), прямої або зворотної — позитивної або негативної, коли змінні варіюють відповідно в однакових або різних напрямках.

У нашому випадку частота сканування, і амплітуда перевищення сигналу файлу зразка змінні — кількісні і рівноцінні у своїх незалежних спостереженнях при їх загальній кількості, тоді емпіричними заходами тісноти їх лінійного взаємозв'язку є коефіцієнт прямої кореляції знаків Г.Т.Фехнера і коефіцієнти парної, приватної та множинної кореляції К.Пірсона.

Коефіцієнт парної кореляції знаків Фехнера визначає узгодженість напрямків в індивідуальних відхиленнях змінних від своїх середніх значень. Він дорівнює відношенню

різниці сум що збігається (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:

$$K_f = \frac{\sum_1^N Mat - \sum_1^N Mit}{\sum_1^N Mat + \sum_1^N Mit} \quad (1)$$

K_f - коефіцієнт Фехнера; Mat (Matching-збігання); Mit (Mitmatching-розбіжність)

Величина K_f - як показано вище, змінюється від -1 до +1 [4,5]. Сумування в (1) проводиться за спостереженнями. Якщо деяке одне відхилення нульове, то воно не входить в розрахунок. Якщо ж відразу обидва відхилення нульові, то такий випадок вважається збігається за знаками і входить до складу. У таблиці 1 наведено статистичні дані для розрахунку коефіцієнта Фехнера.

Таблиця 1.

Номер перевірки	Кількість сканування	Виявлення перевищення амплітуди	Відхилення від середніх		Порівняння	
			x=41	y=43	Mat	Mit
N	x	y	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

Слід зазначити що дані в таблицю 1., заносилися на основі реальних перевірок приміщень на наявність ЗНОІ.

З розрахунків випливає, що коефіцієнт Фехнера дорівнює $K_f = 0,33$

Згідно зі шкалою англійської статистика Чеддока залежність у нас вийшла помірна, що повністю відповідає практичних результатів, кількість сканувань частотного діапазону жорстко не впливає на виявлення сигналу перевищує по амплітуді файл зразка тобто безпосередньо не впливає на виявлення ЗНОІ.

Якщо при аналізі необхідно враховувати та величину відхилення змінних, тоді слід застосовувати коефіцієнти парної, приватної та множинної лінійної кореляції Пірсона [6]. Для їх розрахунку використовують різні методи. Ми будемо використовувати метод прямого рахунку, коефіцієнт парної кореляції Пірсона буде мати вигляд:

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

Для коректного порівняння беремо дані початкові з таблиці 1, проводимо розрахунки, результати попередніх розрахунків та початкових умов представлені в таблиці 2.

Таблиця 2.

Номер перевірки	Кількість сканування	Виявлен. перевищнн амплітуд	Розрахунки		
			x^2	y^2	xy
N	x	y	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	$\sum x^2 = 17500$	$\sum y^2 = 18666$	$\sum xy = 7795$

Провівши розрахунки, данні приведено у таблиці 2, додаючи для аналізу величину відхилення тобто вирахуємо коефіцієнт парної кореляції Пірсона використовуючи при цьому формулу (3). [7,8]. Отримуємо $r_{xy} = 0,4735$, що відповідно до шкали Чеддока відповідає помірної залежності, що повністю підтверджує, результати, раніше проведені в даній статті, для оцінки залежності досліджуваних параметрів за допомогою коефіцієнта Фехнера.

При наявності декількох змінних розраховується коефіцієнт множинної (сукупної) лінійної кореляції Пірсона. Для трьох змінних x, y, z він має вигляд

$$r_{xyz} = \frac{r_{xy} - r_{xz}r_{yz}}{\sqrt{(1 - r_{xz}^2)(1 - r_{yz}^2)}}. \quad (3)$$

У вираз (3), квадрати коефіцієнтів кореляції (5) називаються коефіцієнтами детермінації - відповідно парної або приватної, множинної детермінації:

$$\begin{aligned} d_{xy} &= r_{xy}^2 = r_{yx}^2 = d_{yx} \\ d_{xyz} &= r_{xyz}^2 = r_{yxz}^2 = d_{yxz} \\ D_{y/xz} &= R_{y/zx}^2 \end{aligned} \quad (4)$$

Ці коефіцієнти оцінюють ступінь варіаційної визначеності в лінійному взаємозв'язку змінних, показуючи частку варіації однієї змінної (y), обумовлену варіацією інший (інших) - x і z . Багатомірний випадок наявності більш як трьох змінних тут не розглядається.

Слід зазначити якщо коефіцієнти Пірсона не підкоряються нормальному закону, то у якості критерію їх ваговитості використовується Z - критерій Фішера. Оцінки Фішера ми частково розглядали при складанні моделі (див вираз (6)), для отримання деякої оцінки, за допомогою якої можна було б стверджувати, що отримане рівняння регресії - статистично надійно. Для цього використовується коефіцієнт детермінації R^2 вираз для окремого випадку описується формулою (5)

Наведені вище розрахунки є вірними тільки для лінійної залежності параметрів, але це не завжди вірно. Ці коефіцієнти неможливо використовувати для нелінійних залежностей між змінними. Однак для перевірки гіпотези, що зв'язок між параметрами розпізнання ЗНОІ можливо рахувати умовно лінійним, обчислимо коефіцієнти залежності для нелінійних значень. Тоб то будьмо розраховувати, що приведенні данні мають нелінійну залежність між собою. Для цього використаємо метод нелінійної кореляції, визначмо коефіцієнт рангової кореляції Спірмена. Коефіцієнт рангової кореляції Спірмена - це кількісна оцінка статистичного вивчення зв'язку між явищами, зв'язок між якими нелінійний (нелінійна кореляція). Коефіцієнт рангової кореляції Спірмена визначається:

$$r = 1 - \frac{6 \sum d^2}{n \cdot (n^2 - 1)} \quad (6)$$

Де $d = d_k - d_m$ різниця між рангами;

n – число рангів.

Вираз (6), показує, як відрізняється отримана при спостереженні сума квадратів різниць між рангами від випадку відсутності зв'язку.

Ранг – це порядковий номер, який призначається значенням, якщо зустрічаються два однакових значення, їм присвоюють однакове значення рангу, рівне середньому арифметичному рангів цих значень.

При рангової кореляції порівнюють не власними значення вимірювань, а тільки порядок (ранги), тому обчислення рангового коефіцієнта можливо тільки тоді, коли результати вимірювань отримані на основі шкали не нижче порядкової. Рангові коефіцієнт не рекомендується застосовувати, якщо пов'язаних пар менш як 5 і більш як 20.

Для визначення коефіцієнта рангової кореляції Спірмена, візьмемо статистичні данні з таблиць 1 та 2. Проведемо розрахунки. Початкові дані та данні розрахунків наведені у табл. 3.

Таблиця 3.

Номер перевірки	Кількість сканування	Виявлення перевищення амплітуди	Розрахунки		
			Ранг X	Ранг Y	$(X - Y)^2$
N	x	y			
1	55	43	1	7	36
2	45	42	2	5,5	12,25
3	35	39	3	2	1
4	25	40	4	3	1
5	35	42	5	5,5	0,25
6	45	47	6	9	9
7	50	53	7	10	9
8	40	38	8	1	49
9	35	45	9	8	1
10	45	41	10	4	36
Разом	410	430	55	55	154,5

Обчислимо значення за виразом (6),

$$p = 1 - \frac{6 \sum d^2}{n \cdot (n^2 - 1)} = 1 - \frac{6 \cdot 154,5}{10(100 - 1)} = 0,06363$$

Коефіцієнт рангової кореляції Спірмена дорівнює 0,06363, що згідно критерія статистика Чеддока слабка — залежність.

Статистична значимість отриманого коефіцієнта, ще може оцінюватися за допомогою t-критерію Стьюдента. Якщо розраховане значення t-критерію менше табличного при заданому числі ступенів свободи, статистична значимість спостерігається взаємозв'язку - відсутній. Якщо більше, то кореляційний зв'язок вважається статистично значущою.

Розрахуємо t-критерій Стьюдента для нашого випадку:

$$t = \frac{p\sqrt{n-2}}{\sqrt{1-p^2}} = \frac{0,06363\sqrt{10-2}}{\sqrt{1-0,06363^2}} = 0,1803$$

За табличним значенням при ймовірності 0,95 цей коефіцієнт складає 2,2615, це значно перевищує значення наших розрахунків. Доведено, що статистична значимість взаємозв'язку - відсутня.

Виходячи з вищевикладеного можна зробити висновок, що розроблений метод оцінки впливу параметрів сигналу один на одного, за допомогою яких розпізнається ЗНОІ, який засновано на основі кореляційно-регресійного аналізу підтверджується двома різними методами оцінки для передбачуваної лінійної залежності, та методом оцінки якщо параметри мають нелінійну залежність, це свідчить про те що параметри які ми визначили для розпізнавання ЗНОІ, незалежні один від одного.

Використавши розроблений метод можливо викреслити на одному з етапів пошуку ЗНОІ параметри які мають сильну залежність один від одного тим самим скоротити час сканування та повисить ймовірність розпізнавання ЗНОІ.

Проводячи перевірку усіх параметрів розпізнавання ЗНОІ на взаємозалежність, ми відібрали основні параметри це:

1. Спектр сигналу;
2. Спекральна енергія сигналу;
3. Фазочастотні характеристики;
4. Амплітуда сигналу;
5. Гармоніки сигналу.

Але викреслювати другі параметри розпізнавання, цілком, неможливо так як ці параметри можуть допомогти визначити ЗНОІ на послідовному етапе.

Так перевищення амплітуди порога (визначного файлу «зразка») на де якої частоті, ще не є остаточним признакам ЗНОІ, а поява цього сигналу на частоті кратної двом та кратної трем у сукупності з спектром сигналу, дає повну гарантію розпізнавання ЗНОІ.

Таким чином визначення взаємозв'язку на основі кореляційно регресійного аналізу параметрів розпізнавання ЗНОІ визначено необхідний перелік параметрів для первинного і потім остаточного розпізнавання ЗНОІ.

Напрямок подальших досліджень є визначення числових оцінок кореляційно-регресійної моделі розпізнавання засобів негласного отримання інформації на предмет її статистичної надійності.

Висновки

Запропоновано метод оцінки взаємовпливу сигналів ЗНОІ на основі кореляційно-регресійного аналізу, для лінійної залежності параметрів розпізнавання ЗНОІ двома різними

методами, які показали практично однакові результати, що свідчить про достовірності результатів і правильності запропонованої методики розрахунків.

За допомогою запропонованого алгоритму визначення коефіцієнта рангової кореляції Спірмена, визначили параметри розпізнавання ЗНОІ з нелінійною залежністю, отримані результати підтверджують відсутність взаємозалежності запропонованих параметрів розпізнавання ЗНОІ.

Обчисленим *t*-критерієм Стьюдента, для коефіцієнта Спірмана, підтвердили адекватність запропонованого методу оцінки взаємовпливу параметрів розпізнавання ЗНОІ.

Проведені розрахунки дозволяють вибрати самі основні параметри розпізнавання, що дозволить на першому етапі скоротити час розпізнавання на 15%.

Перелік посилань

1. Регрессионный анализ [Электронный ресурс]. Режим доступа: <http://www.machinelearning.ru/> (01.05.2019).
2. Методы изучения взаимосвязи социально-экономических явлений с помощью корреляционно-регрессивного анализа [Электронный ресурс]. Режим доступа: www.Grandars.ru » Статистика » Общая теория статистики » (01.07.2019).
3. Laptev A.A. Barabash O.V., Savchenko V.V., Savchenko V.A., Sobchuk V.V. The method of searching for digital means of illegal reception of information in information systems in the working range of Wi-Fi /International Journal of Advanced Research in Science, Engineering and Technology. India (ISSN: 2350-0328) 2019. Vol. 6, Issue 7. – P. 10101-10105
4. Laptev A.A. Sachenko V.A., Barabash O.V., Sachenko V.V., Matsko A.I The metod of searching for digital vtfns of illegal obtaning of information on the basis cluster analysis// Magyar Tudományos Journal. Budapest, Hungary (ISSN 1748-7110) 2019. № 31 P33-37
5. Радченко, С. Г. Методология регрессионного анализа : [монография] / С. Г. Радченко. - Киев : Корнійчук, (ISBN 978-966-7599-72-0) 2011. - 375 с.
6. Петров Ю.П. Как получать надежные системы уравнений/Ю.П.Петров –СПб:БХВ-Петербург,2009-176 с.
7. Методы корреляционного и регрессионного анализа [Электронный ресурс]. Режим доступа: <https://studme.org/33796/informatika/metody/> / (07.07.2019).
8. Методы сравнительной комплексной оценки хозяйственно-финансовой деятельности организации [Электронный ресурс]. Режим доступа: / <https://lesnaya.nethouse.ru/articles/33525/> (27.07.2019).

Надійшла: 30.07.2019

Рецензент: д.т.н., проф. Вишнівський В.В.