DOI: 10.18372/2225-5036.29.17551

ВИБІР ПОКАЗНИКІВ ПРОГНОЗУВАННЯ КІБЕРЗАХИЩЕНОСТІ КОМП'ЮТЕРНИХ СИСТЕМ

Володимир Хорошко, Юлія Хохлачова, Наталія Вишневська

Національний авіаційний університет

ХОРОШКО Володимир Олексійович, д.т.н., професор.

Рік та місце народження: 1945 рік, м. Харків, Україна.
Освіта: Київський інститут інженерів цивільної авіації, 1968 рік.
Посада: професор кафедри безпеки інформаційних технологій.
Наукові інтереси: інформаційна безпека, технічні системи захисту інформації, аналіз функіонування складних систем.
Публікації: більше 500 наукових публікацій, серед яких наукові статті, монографії, підручники та навчально-методичні посібники.
E-mail: professor_va@ukr.net.
ORCID ID: 0000-0001-6213-7086.

ХОХЛАЧОВА Юлія Євгеніївна, к.т.н., доцент.
Рік та місце народження: 1981 рік, м. Київ, Україна.
Освіта: Національний авіаційний університет, 2004 рік.
Посада: доцент кафедри безпеки інформаційних технологій.
Наукові інтереси: інформаційна безпека, оцінювання уразливостей, оптимізація інформаційних систем.
Публікації: більше 100 наукових публікацій, серед яких наукові статті, монографії, підручники та навчально-методичні посібники.
E-mail: yuliiahohlachova@gmail.com.
ORCID ID: 0000-0002-1883-8704.

ВИШНЕВСЬКА Наталія Сергіївна, старший викладач кафедри безпеки інформаційних технологій Національного авіаційного університету.

Рік та місце народження: 1977 рік, м. Київ, Україна.
Освіта: Національний авіаційний університет, 2001 рік.
Наукові інтереси: кіберзахищеність інформаційних систем, кіберризики енергетичних систем.
E-mail: nataliia.vyshnevska@npp.nau.edu.ua.
ORCID ID: 0000-0001-9036-6556.

Анотація. *В статті запропонований алгоритм вибору показників прогнозування кіберзахищенності комп'ютерних систем. Процеси кіберзахисту відносяться до випадкових багатовимірних, динамічних нестаціонарних, активних (цілеспрямованих), що ускладнює завдання прогнозування показників кіберзахищенності. Аналіз публікацій показав складність вибору найефективнішого методу прогнозування кіберзахищенності, який полягає у визначенні щодо класифікації методів прогнозування характеристик кожного методу, переліку вимог до ретроспективної інформації. Таким чином, застосування екстраполяції у прогнозуванні завжди передбачає використання будь-яких моделей, тому моделювання є основою для екстраполяції. Прогнозування є досить складним завданням, що підтверджується аналізом причин та факторів, які потенційно впливають на зміни прогнозованого показника. Вирішення такого завдання, як і будь-якого іншого складного завдання, потребує системного підходу, який допомагає зрозуміти суть проблеми та вибрати адекватні методи його вирішення, а також оцінить причини можливих невдач. Отриманий алгоритм містить багатоповерховість моделі, як у класі лінійних, так і в класі нелінійних за вхідними змінними моделями; виключення окремих членів кращого приватного опису та на основі цього розширення базисного набору аргументів; є оптимальним*

за обчислювальними витратами для ітераційних алгоритмів методу групового обчислювального алгоритму схеми розрахунку критерію іспиту, що ковзає. А також має можливість оцінювати коефіцієнти у моделях як за методом найменших квадратів, так і за методом найменших модулів.

Ключові слова: *кіберзахищеність, комп'ютерні системи, показники прогнозування, методи прогнозування, алгоритми прогнозування.*

Постановка проблеми

Процеси кіберзахисту відносяться до випадкових багатовимірних, динамічних нестаціонарних, активних (цілеспрямованих), що ускладнює завдання прогнозування показників кіберзахищеності.

При встановленні якості прогнозів треба пам'ятати таке. Оцінки ймовірностей подій, що даються після того, як подія відбулася, помітно відрізняються від оцінок, що даються до здійснення події, оскільки люди схильні завершувати ймовірність того, що сталося.

Тому причини того, що прогнози не справджуються, можна сформулювати так:

- неправильний вибір методу прогнозування;
- помилки у використанні методу прогнозування;
- недостатність або помилковість вихідних даних;
- недостатні витрати інтелектуальних сил на складання прогнозів.

Компонентами системи прогнозування є такі:

1) теоретичні засади;
2) технології прогнозування;
3) вихідні відомості;
4) колекції чужих прогнозів;
5) експерти у предметних галузях;
6) фахівці з прогнозування:
- розробники теорії;
- розробники технологій;
- користувацькі технології;
- засоби обробки даних.

Якщо об'єкт прогнозування добре відомий, то основна складність прогнозування полягає у переробці великої кількості наявних відомостей, а якщо про нього мало що відомо, то основна складність – у нестачі відомостей.

Якщо об'єкт прогнозування не ізольований, а зазнає істотного впливу ззовні, виникає необхідність прогнозувати не тільки процеси в самому об'єкті, але й вплив на об'єкт.

Система прогнозування не може працювати весь час з тим самим горизонтом, з одним і тим же ступенем достовірності: бувають нестійкі ситуації, в яких навіть невелика перевага деяких впливів у довільну сторону знижує точність прогнозу.

За ступенем формалізації методи прогнозування можна поділити на інтуїтивні та формалізовані.

Інтуїтивні методи прогнозування використовують оцінки експертів. Розрізняють індивідуальні та колективні експертні оцінки.

До групи формалізованих методів входять методи екстраполяції (найменших квадратів, експонентційного згладжування, ковзних середніх, адаптивного згладжування) та моделювання (структурного, мережевого, матричного, імітаційного).

При такому поділі методам екстраполяції протиставляються як самостійні методи моделювання.

З одного боку, побудова моделей має на меті розкрити закономірність розвитку об'єкта, що вивчається, або процесу на деякій ретроспективній ділянці. Якщо модель адекватно відбиває зв'язку та властивості об'єкта, вона може бути основою для екстраполяції, тобто. перенесення деяких висновків щодо поведінки моделі на об'єкт. З іншого боку, методи екстраполяції – не що інше, як використання теоретичних та емпіричних моделей для знаходження змінних поза ретроспективною участю спостережень за даними залежностей між ними на ретроспективній ділянці.

Таким чином, застосування екстраполяції у прогнозуванні завжди передбачає використання будь-яких моделей, тому моделювання є основою для екстраполяції.

Аналіз останніх досліджень і публікацій

Аналіз публікацій, присвячених прогнозуванню, дозволяє зазначити, що відома велика кількість методів прогнозування [1, 2, 3].

При прогнозуванні треба враховувати різницю між ймовірністю і можливістю. Імовірність - характеристика подій, що багато разів повторюються, а можливість - характеристика унікальних і рідкісних.

Способи прогнозування:

- на основі встановлених закономірностей;
- за аналогією;
- екстраполяцій тенденцій;
- через з'ясування схильностей, намірів та можливостей суб'єктів, які визначають події або помітно впливають на них.

Мета і постановка завдання

Складність вибору найефективнішого методу прогнозування кіберзахищеності полягає у визначенні щодо класифікації методів прогнозування характеристик кожного методу, переліку вимог до ретроспективної інформації.

Прогнозування є досить складне завдання, що підтверджується аналізом причин та факторів, які потенційно впливають на зміни прогнозованого показника. Тому метою є вирішення такого завдання, як і будь-якої іншої складної задачі, що потребує системного підходу, який допомагає зрозуміти суть проблеми та вибрати адекватні методи її вирішення, а також оцінить причини можливих невдач.

Виклад основного матеріалу

Вибір підходу до прогнозування показників кіберзахищеності.

Прогнозування кіберзахищеності об'єктів є одним із вирішальних наукових факторів формування стратегії та тактики кіберзахисту.

Для прогнозування та моделювання процесів кіберзахисту найбільш прийнятними є статистичні методи, що ґрунтуються на існуючих тенденціях у змінах показників кіберзахищеності.

Моделі прогнозування можуть бути як довгостроковими, так і короткостроковими. Внаслідок широкого ступеня невизначеності в отриманні інформації пріоритетне значення надається короткостроковим прогнозам.

Невід'ємною складовою завдань управління (оптимального, автоматизованого чи на рівні прийняття рішень) є побудова моделей, які описують чи прогнозують поведінку об'єкта, процесу системи. У випадку отримання математичної моделі необхідно вибрати її структуру і оцінити параметри, тобто розв'язати задачу структурної ідентифікації.

Структурна ідентифікація сприймається, як завдання перешкоди структури моделі з мінімальною дисперсією помилки прогнозування.

У загальному випадку процес вирішення задачі структурно-параметричної ідентифікації включає наступні основні етапи:

- завдання вибірки даних до апріорної інформації;
- вибір чи завдання класу базисних функцій та перетворення даних;
- генерація різних структур у вибраному класі;
- оцінювання параметрів генерованих структур та формування безлічі моделей F;
- мінімізація заданого критерію та вибір оптимальної моделі f;
- перевірка адекватності отриманої оптимальної моделі;
- ухвалення рішення про завершення процесу.

Перелічені етапи описують довільний процес побудови моделей, причому залежно від апріорної інформації та мети моделювання ті чи інші етапи можуть бути відсутніми.

Зазначимо, що завдання моделювання та прогнозування тісно пов'язані між собою, але не тотожні. Завданням моделювання є побудова математичних моделей для кількісного опису зв'язку між вхідними показниками (залежними змінними) процесів, що моделюються, і вхідними (незалежними) змінними.

Прогнозування - кількісне передбачення майбутнього розвитку процесу за моделями, отриманими тим чи іншим методом. При цьому завдання прогнозування може виступати як екстраполяція закономірного розвитку процесів, що моделюються, на майбутній проміжок часу або як імітаційне моделювання, тобто. застосування математичних моделей на дослідження зміни модельованих процесів з урахуванням різних варіантів поведінки незалежних змінних.

Існують різноманітні постановки завдання структурної ідентифікації, методи її вирішення та варіанти програмної реалізації, розроблені фахівцями у галузях ідентифікації, прикладного регресійного аналізу та пошуку залежностей [1, 4, 5, 6].

Більшість методів побудовано різних підходах, що ускладнює порівняльний аналіз визначення умов їх ефективного застосування. Оскільки для завдань управління важливо будувати моделі з меншою помилкою прогнозування, це є основою для порівняння ефективності існуючих підходів.

Найбільш поширеними є дві групи методів, на основі яких будують моделі для прогнозування:

Описання взаємодії процесів кіберзахисту різного рівня на основі вивчення внутрішніх механізмів функціонування процесів, що моделюються, та представлення отриманих знань;

Отримання моделей за допомогою оцінювання параметрів за вибірками спостережень із застосуванням регресійного аналізу;

Ці групи методів відносяться до дедуктивних, які реалізують принцип від загального до приватного.

На відміну від них індуктивні методи моделювання базуються на принципі від приватного до загального, тобто. від емпіричних даних про конкретні процеси до побудови моделей за допомогою інформаційних технологій отримання знань із даних.

При цьому математичні моделі будуються за вибірками спостережень за допомогою автоматичного генерування та порівняння великої кількості різних моделей з вибором найкращої з позиції відповідності інформації, яка неявно міститься в даних.

Мета підходу – виявлення прихованих закономірностей, тобто, пошуку опису закономірних причинно-наслідкових зв'язків між елементами досліджуваних процесів на основі статистичних даних, у яких ці зв'язки об'єктивно відображені.

З погляду завдань управління процесами кіберзахисту об'єктів доцільно прагнути отримати в результаті ідентифікації модель з найкращими властивостями, що прогнозують.

У зв'язку з цим із усіх підходів, що існують у структурній ідентифікації, найбільший інтерес представляє «перехресне підтвердження». За своєю суттю він спрямований на вибір моделей, що зменшують помилку прогнозування: у найпростішому випадку вибірка ділиться на дві частини, на одній з яких генерується безліч моделей, а на другій обчислюється помилка кожної моделі, і вибирається та, яка краще працює «в режимі прогнозування» .

Серед методів цього напряму виділяється метод групового обліку аргументів (МГОА) [1, 7] як метод автоматичного пошуку кращої моделі, що має велику

різноманітність можливостей порівняно з іншими методами. Це індуктивний метод побудови лінійних, нелінійних, різницевих та інших математичних моделей складних процесів із коротких вибірок даних. В його основу покладено принципи зовнішнього доповнення, автоматичної генерації та послідовної селекції структур моделей, що ускладнюються.

Моделі побудовані за алгоритмами МГОА, за своїми властивостями, що прогнозують, значно перевершують регресійні в силу того, що за цими алгоритмами автоматично відбираються аргументи (фактори), найбільш інформативні для даного об'єкта моделювання.

Для складних об'єктів кіберзахисту та процесів типовою є невизначеність інформації щодо механізму їх функціонування, ступеня інформативності вимірювальних змінних та властивостей. При цьому необхідно застосовувати формалізовані методи та автоматизовані способи моделювання.

Тому дуже актуальною є проблема розробки алгоритму структурної ідентифікації прогнозуючих моделей з метою створення автоматизованих способів оптимального вибору структури моделей складних об'єктів за вибірками обмеженого обсягу в умовах неповної інформації.

Побудова лінійних та нелінійних моделей на основі МГОУ

Розглядається статистичний об'єкт з входами m і одним виходом, результати спостережень за яким представлені матрицею $X[Nxm]$ і вектором $Y[Nx]$ При цьому вважатимемо, що входи – це кібератаки, а вихідні величини – реакція на них.

За даними N спостережень потрібно визначити структуру залежності одномірної вхідної змінної Y від набору вхідних змінних X в умовах, коли апріорно невідомо, які саме змінні з набору вхідних змінних беруть участь у формуванні вихідної змінної, а також невідома дисперсія помилки у спостереженнях вихідної змінної Y.

Нехай вихідна модель об'єкта може належати деякій невідомій множині C, в якій кожна модель має вигляд:

$$\bar{y} = f(X, \overline{\theta_f}), \qquad (1)$$

де $\overline{\theta_f}$ – вектор параметрів моделі, що оцінюється тим чи іншим способом за даними спостережень.

Завдання зводиться до пошуку мінімуму заданого критерію J якості моделі:

$$f^* = \arg\min_{f \in G} J\left[y, f(x,\theta)\right]. \qquad (2)$$

Ця постановка завдання показує, що це методи її вирішення можуть відрізнятися, по крайнього заходу, за такими признаками:

- методу формування (генерації) структур моделей з множини G;
- методу оцінювання параметрів цих моделей;
- самим критерієм J оцінки їхньої якості;
- організацією руху до мінімуму J.

Методи оцінювання параметрів моделей, критерії їх якості та методи пошуку мінімуму критеріїв в основному незалежні між собою та можуть застосовуватись у різних поєднаннях. Тому може бути запропоновано безліч різних методів розв'язання задач (1)-(2) [8, 9].

Розглянемо це завдання відповідно до ідеології методу групового обліку аргументу (МГОА). Доцільність використання МГОА пояснюється лише тим, що у ньому реалізуються ітераційні схеми ускладнення моделей. Ускладнення моделей ряду до ряду селекції відбувається за рахунок схрещування кращих моделей попереднього ряду.

Основні засади МГОА:

- проміжне рішення не єдине;
- зовнішнє доповнення (використання перевірочної вибірки);
- саме добір проміжних рішень;
- єдність остаточного рішення.

Уявлення про найкращу структуру моделі (1) досить умовно залежить від вимог дослідника. Зазвичай припускають, що Y можна апроксимувати поліномами. Якщо ступінь полінома не вище k, то завдання вибору структури моделі можна у вигляді пошуку функції $g_1(V,k)$ вибору змінних завдань на вершинах одиничного гіперкуба $[0,1]^m$. З усіх вершин потрібно знайти таку, в якій досягається екстремум $g_1[V,k): g_r(V,k) \to \min$, де r число вибраних змінних, $1 \le r \le m$, $V = (v_1, v_2, v_3, ...v_m)$, $v_j = 1$, якщо змінна x_j входить у модель, $v_j = 0$ – інакше.

При виборі структури моделі керуватимемося двома принципами:

- прагнути до того, щоб модель була простішою, тобто включала якнайменше коефіцієнтів полінома (принцип економічності);
- покращувати модель, перевіряючи її адекватність (принцип адекватності).

Для мінімізації заданого критерію (2) визначеного на безлічі моделей G застосовуються як перебірні, так і ітераційні алгоритми.

Аналіз перерахованих критеріїв та методів показав, що для побудови математичної моделі нелінійної регресійної залежності виду (1)-(2) доцільно використовувати:

-МГОА як метод перебору моделей;

-метод найменших квадратів та метод найменших модулів як методи оцінки параметрів моделі;

-критерії залишкової суми квадратів, регулярності, ковзного контролю з метою оцінки якості отриманих моделей.

Нехай закон функціонування досліджуваного об'єкта має вигляд:

$$y = \overset{0}{y} + \xi = \sum_{j=1}^{\overset{0}{m}} \overset{0}{\theta}_j \cdot \overset{0}{x}_j + \xi \ , \qquad (3)$$

де y - Вихід об'єкта, що спостерігається; $\overset{0}{y}$ – незахищений вихід об'єкта, що не спостерігається; ξ – випадкова помилка вимірювання, що не спостерігається; $\overset{0}{x}_j$ – j –й вхід об'єкта з безлічі входів $\overset{0}{x}$, що беруть участь у формуванні виходу об'єкта; $\overset{0}{m}$ - число входів, що належать безлічі $\overset{0}{x}$; $\overset{0}{\theta} = (\overset{0}{\theta}_1, \overset{0}{\theta}_2, \overset{0}{\theta}_3, ..., \overset{0}{\theta}_{\overset{0}{m}})$ – вектор не рівних нулю невідомих коефіцієнтів.

Безліч входів $\overset{0}{x}$ невідомо. Відомо лише, що $\overset{0}{x} \subset x$, де x – деяка безліч точно вимірюваних m входів об'єкта.

Нехай у результаті спостереження за об'єктом отримано: x – матриця (розмірності Nxm) N спостережень m входів множини x, $rang\, x = m$; y – вектор (розмірності N) відповідних спостережень вихідної величини y.

Відповідно до закону функціонування об'єкта (3) має виконуватись рівність:

$$y = y^0 + \xi = \overset{0}{X}\,\overset{0}{\theta} + \xi \, , \qquad (4)$$

де y^0 - вектор (розмірності N) значень не спостережуваного та не зашумленого виходу об'єкта; $\overset{0}{X}$ - матриця (розмірності $Nx\overset{0}{m}$) спостережень входів об'єкта, що належать множині $\overset{0}{x}$; ξ - вектор (розмірності N) випадкових помилок вимірювання, що спостерігаються.

Нехай щодо ξ виконується таке припущення: $E\{\xi\} = O_N$, $E\{\xi\xi^T\} = \delta^2 I_N$, де $E\{\cdot\}$ - знак математичного очікування; T – знак транспонування; O_N – нульовий вектор (розмірності $N \times 1$); δ^2 – невідома кінцева величина; I_N - поодинока матриця (розмірності $N \times N$).

Виходячи з цього потрібно знайти:

-безліч $\overset{0}{x}$;

-оцінку вектора коефіцієнтів $\overset{0}{\theta}$;

-оцінку дисперсії помилок вимірів δ^2.

Приймемо, що клас моделей, що синтезуються, має вигляд:

$$\hat{y} = \sum_{q=1}^{s} \theta_q \cdot \prod_{j=1}^{m} x_j^{\alpha_{jq}} \ , \qquad (5)$$

де $\hat{y}$ - вихідна змінна; s – число членів моделі; θ_q, $q = 1, 2, 3, ..., s$ – коефіцієнти; x_j, $j = 1, 2, 3, ..., m$ – вхідні змінні; m – число вхідних змінних; α_{jq} - показник ступеня, в якій змінна x_j входить в q-й член.

Частним описом називатимемо вектор $\bar{z}$ (розмірності N), отриманий на деякій ітерації алгоритму як наближення вектора $\bar{y}$.

Структурною частиною опису називаємо набір параметрів α_{jq} і s, що визначають $\bar{z}$ у поданні (5).

Для побудови ітераційного алгоритму МГОА необхідно:

- вказати початкову матрицю приватних описів $\bar{z}^0$;
- визначити оператор R, який здійснює відображення, $\bar{Z}^{r-1} \xrightarrow{R} \bar{Z}^r$, $r = 1, 2, 3, ...,$ – номер ітерації;
- вказати правило завершення ітерацій.

Загальний вигляд матриці приватних описів $\bar{Z}$ визначаємо так:

$$\bar{Z}^r = \left[\bar{z}_1^r : \bar{z}_2^r : \bar{z}_3^r : ... : \bar{z}_{F+2+m+2s}^r \right], \qquad (6)$$

де $\bar{z}_1^r$, $j = 1, 2, 3, ... F + 2 + m + 2s$ – вектори (розмірності N), частний опис; F – число кращих приватних описів, що передаються від ітерації до ітерації; S – число членів у структурі кращого приватного опису ітерації $(r-1)$.

Позначимо:

$$\bar{G}^r = \left[\bar{z}_1^r : \bar{z}_2^r : \bar{z}_3^r : ... : \bar{z}_F^r \right];$$

$$\bar{C}^r = \left[\bar{z}_{F+1}^r : \bar{z}_{F+2}^r : ... : \bar{z}_{F+2+m}^r \right]; \qquad (7)$$

$$\bar{D}^r = \left[\bar{z}_{F+2+m+1}^r : \bar{z}_{F+2+m+2}^r : ... : \bar{z}_{F+2+m+2s}^r \right].$$

Алгоритм складається з наступних кроків:

Крок 1. Вказується початкова матриця приватних описів Z^0 (для неї вважають $s = 0$):

$$Z^0 = [O : o : I : X] = \left[O : \bar{C}^0 \right], \qquad (8)$$

де o – нульовий вектор (розмірності N); O – нульова матриця (розмірності $N \times F$); I – одиничний вектор (розмірності N); $X = [x_1 \vdots x_2 \vdots x_3 \vdots ... x_m \vdots]$ – матриця спостережень вхідних змінних (розмірності $N \times F$).

Крок 2. Визначається вектор R.

Нехай вектори $\bar{z}^r$ будуються за правилом:

$$\bar{z}^r(i) = \bar{a} \cdot \bar{z}^r_{j_1}(i) + \bar{b}\bar{z}^{r-1}_{j2}(i) \cdot \bar{Z}^{r-1}_{j3}(i), \quad (9)$$

де $z = 1, 2, 3, ...$ – номер ітерації; $i = 1, 2, 3, ..., N$ – номер спостереження, який для зручності запису взято у дужки; $j_1, j_2, j_3 = 1, 2, 3, ..., F + 2 + m + 2s, (j_3 \geq j_2)$ – номери приватних описів із матриці $\bar{Z}^{r-1}$; $\bar{a}, \bar{b}$ - коефіцієнти, що визначаються на навчальній підвибірці спостережень (A).

Значення коефіцієнтів ($\bar{a}$) і ($\bar{b}$) перебувають як розв'язання задачі мінімізації:

$$\bar{a}, \bar{b} = \arg\min_{a,b} \Phi(a,b), \Phi(a,b) = \sum_{i=1}^{N(A)} e^2_A(i), \quad (10)$$

де $e_A(i)$, $i = 1, 2, ..., N(A)$ – залишки в регресії y_A за двома змінними:

$$y_A(i) = a \cdot \bar{Z}^{r-1}_{j_1}(i) + b \cdot \bar{Z}^{r-1}_{j_2}(i) \cdot \bar{Z}^{r-1}_{j_3 A}(i) + e_A(i) \quad (11)$$

З усіх генерованих за правилом (9)-(11) приватних описів відбираються F описів кращих мінімуму квадратичної норми залишків на перевірочній підвибірці спостережень B:

$$J = \frac{1}{N(B)} \sum_{i=1}^{N(B)} \bar{e}^r_B(i),$$
$$\bar{e}_B(i) = y_B(i) - \bar{Z}^r_B, \quad (12)$$
$$\bar{Z}^r_B = \bar{a} \cdot \bar{Z}^{r-1}_{j_1 B}(i) + b \cdot \bar{Z}^{r-1}_{j_2 B}(i) \cdot \bar{Z}^{r-1}_{j_3 B}(i),$$

де $N(B)$ – обсяг перевірочної добірки B.

Відібрані найкращі приватні описи, ранжировані за зменшенням величини J, використовуються при формуванні матриці:

$$\bar{G}^r = \left[\bar{z}^r_1 \vdots \bar{z}^r_2 \vdots ... \bar{z}^r_F \vdots \right]. \quad (13)$$

Матриця $\bar{C}^r$ не змінюється $\bar{C}^r = \bar{C}^{r-1}$.

Матриця $\bar{D}^r$ формується з урахуванням структури кращого із F відібраних приватних описів ($\bar{Z}^r_F$). Перші S стовпці матриці $\bar{D}^r$ заповнюються окремими елементами кращого приватного опису за правилом:

$$\bar{d}^r_h(i) = \bar{Z}^r_{F+2+m+2s}(i) = \theta_h \prod_{j=1}^{m} x_j^{\alpha_{hj}}(i), \quad (14)$$

де $h = 1, 2, 3, ..., s$ – номер члена у структурі; s - число членів у структурі кращого приватного опису.

Другі s стовпців матриці $\bar{D}^r$ формуються почерговим винятком окремих членів із структури приватного опису за правилом:

$$\bar{d}^r_{s+h}(i) = \bar{Z}^r_{F+2+m+s+h}(i) = \sum_{\substack{q=1 \\ (q \neq h)}}^{s} \bar{\theta}_q \prod_{j=1}^{m} x_j^{d_{jq}}(i). \quad (15)$$

Таким чином визначається оператор R перетворення $\bar{Z}^{r-1} \xrightarrow{R} Z^r$.

Правило обстановки для ітераційної схеми: обчислення закінчуються на ітерації r^*, якщо виконується умова:

$$J(\bar{Z}^{r^*-1}_F) - J(\bar{Z}^{r^*}_F) < \delta_z, \quad (16)$$

де $J(\bar{Z}^{r^*}_F)$ – значення критерію для найкращого приватного опису ітерації r; δ_z – задане число.

Особливістю алгоритму є багатоповерховість. Номер поточного етапу визначає максимально можливу кількість членів у моделях. Синтез моделей починається з етапу з номером $p = 1$ чи з будь-якого заданого номера p^0.

Кожен етап є ітераційною схемою (8)-(16). Початкова матриця приватних описів етапу з номером p задається кінцевою матрицею приватних описів попереднього етапу:

$$\bar{Z}^0_p = \bar{Z}^{r^*}_{p-1}, \quad (17)$$

а для $p = p^0$ вона збігається з (8). Обчислення закінчується на етапі p^*, якщо виконано умову:

$$J(\bar{Z}^{r^*}_{F_1 p^*-1}) - J(\bar{Z}^{r^*}_{F_1 p^*}) < \delta_p \quad (18)$$

де $J(\bar{Z}^{r^*}_{F_1 p^*})$ – значення критерію для кращого частного опису r-й ітерації етапу p; δ_p – задане число.

Висновки. Відмінні риси алгоритму:

- багатоповерховість моделі;
- пошук моделі як у класі лінійних, так і в класі нелінійних за вхідними змінними моделями;
- прийоми виключення окремих членів кращого приватного опису та на основі цього розширення базисного набору аргументів;

- оптимальність за обчислювальними витратами для ітераційних алгоритмів МГОА схеми розрахунку критерію іспиту, що ковзає;
- можливість оцінювати коефіцієнти у моделях як за методом найменших квадратів, так і за методом найменших модулів.

Список літератури

[1] Івахненко О.Г., Лапа В.Г. Передбачення випадкових процесів К: Наук. думка, 1981. 216 с.

[2] Згуровський М.З., Панкратова Н.Д. Технологическое предвидение. К: Политехника, 2005. 165с.

[3] Івахненко О.Г. Довгострокове прогнозування та управління складними системами. К: Техніка, 1975. 264 с.

[4] Івахненко О.Г. Юрачковський Ю.П. Моделювання складних систем за екстримальними даними. К: Техніка, 1989. 121 с.

[5] Опірський І.Р. Загальні проблеми прогнозування НСД в інформаційних системах держави // Правове, нормативне та метрологічне забезпечення системи захисту інформації в Україні. Вип 2(30). 2015. С. 31-34.

[6] Опірський І.Р. Особливості прогнозування несанкціонованого доступу при недостатній апріорній інформації // Інформаційна безпека. №1(13). 2014. С. 5-11.

[7] Томашевський В.М. Моделювання систем. К: Вид. група BHV, 2007. 352 с.

[8] Вучков І., Бояджиєва Л., Солаков Є. Прикладний лінійний регресійний аналіз. Вид. 2-е / Пер. з болгарського. М: Фінанси та кредит, 1997. 239 с.

[9] Себер Дж. Лінійний регресійний аналіз. М: Мир, 1980. 456 с.

УДК 004.681.3

Khoroshko V., Khokhlachova Y., Vyshnevska N. Choice of indicators for forecasting cyber protection of computer systems

Abstract. *The article proposes an algorithm for selecting indicators for predicting cyber security of computer systems. Cyber defense processes are random multidimensional, dynamic non-stationary, active (targeted), which complicates the task of forecasting cyber security indicators. The analysis of publications showed the complexity of choosing the most effective method of forecasting cyber security, which consists in determining the classification of methods of forecasting the characteristics of each method, the list of requirements for retrospective information. Thus, the use of extrapolation in forecasting always involves the use of any models, so modeling is the basis for extrapolation. Forecasting is a rather complex task, which is confirmed by the analysis of the causes and factors that potentially affect changes in the forecasted indicator. Solving such a task, like any other complex task, requires a systematic approach that helps to understand the essence of the problem and choose adequate methods of solving it, as well as assess the reasons for possible failures. The resulting algorithm contains a multi-story model, both in the class of linear and in the class of nonlinear models according to input variables; exclusion of individual members of a better private description and based on this expansion of the basic set of arguments; is optimal in terms of computational costs for the iterative algorithms of the method of the group computational algorithm of the sliding test criterion calculation scheme. And it also has the ability to estimate coefficients in models both by the method of least squares and by the method of least modules.*

Keywords: *cyber security, computer systems, forecasting indicators, forecasting methods, forecasting algorithms.*

Хорошко Володимир Олексійович, доктор технічних наук, професор, професор кафедри безпеки інформаційних технологій Національного авіаційного університету.
Volodymyr Khoroshko, doctor of technical sciences, professor, professor of the department of security of information technologies of the National Aviation University.

Хохлачова Юлія Євгеніївна, кандидат технічних наук, доцент, доцент кафедри безпеки інформаційних технологій Національного авіаційного університету.
Yuliia Khokhlachova, candidate of technical sciences, associate professor, associate professor of the department of security of information technologies of the National Aviation University.

Вишневська Наталія Сергіївна, старший викладач кафедри безпеки інформаційних технологій Національного авіаційного університету.
Nataliya Vyshnevska, senior lecturer at the Information Technology Security Department of the National Aviation University.

Отримано *6 березня 2023 року*, затверджено редколегією *27 березня 2023 року*