

УДК 004.056.53

доктор філософії Фесьоха В. В. ORCID: 0000-0001-6612-1970 (ВІТІ ім. Героїв Крут)
Фесьоха Н. О. ORCID: 0000-0002-9797-5589 (ВІТІ ім. Героїв Крут)

МЕТОД РЕГУЛЯРИЗАЦІЇ ОЗНАКОВОГО ПРОСТОРУ БІОМЕТРИЧНОЇ МОДЕЛІ КЛАВІАТУРНОГО ПОЧЕРКУ КОРИСТУВАЧІВ ІНФОРМАЦІЙНИХ СИСТЕМ ВІЙСЬКОВОГО ПРИЗНАЧЕННЯ НА ОСНОВІ ФАКТОРНОГО АНАЛІЗУ

У статті вирішується актуальне наукове завдання регуляризації ознакового простору біометричної моделі клавіатурного почерку користувачів інформаційних систем військового призначення з метою підвищення ефективності процедури автентифікації користувачів системами контролю і розмежування доступу. Постановка даного наукового завдання зумовлена наявністю наступних недоліків існуючих біометричних моделей клавіатурного почерку користувачів: при збільшенні числа користувачів у системі зростає складність їх класифікації; складність формалізації унікальності користувачів; надто складна реалізація визначення факту підміни уже авторизованого користувача; ознаковий простір існуючих біометричних моделей є незначним в силу обмеження множини властивостей сучасної клавіатури, що негативно впливає на показник точності процедури автентифікації. Обрано біометричну модель клавіатурного почерку, особливістю якої є формалізація унікальності користувача інформаційної системи на основі виявлення властивих йому закономірностей клавіатурного почерку нечіткими правилами. Запропоновано удосконалений метод регуляризації ознакового простору біометричної моделі клавіатурного почерку користувачів інформаційних систем військового призначення. Суть запропонованого методу, яка відрізняє його від існуючих, полягає у тому, що збільшення множини ознак біометричної моделі досягається шляхом додавання до неї виявлених прихованих фактів із множини власних ознак на основі факторного аналізу із найбільшим показником їх мінливості. Застосування запропонованого методу дозволяє вирішити завдання нелінійної сепарабельності n -вимірного ознакового простору біометричної моделі клавіатурного почерку, що у свою чергу дозволяє підвищити показники точності та достовірності процедури автентифікації системами контролю і розмежування доступу інформаційних систем військового призначення.

Ключові слова: регуляризація, біометрична модель, поведінкова біометрія, клавіатурний почерк, інформаційні системи, автентифікація, несанкціонований доступ, факторний аналіз.

V. Fesokha, N. Fesokha *The method of regularizing the sign space of the biometric model of the keyboard handwriting of users of military information systems on the basis of factor analysis.*

The article addresses the current scientific task of regularizing the feature space of the biometric model of the keyboard handwriting of users of military information systems in order to increase the effectiveness of the user authentication procedure by access control and demarcation systems. The setting of this scientific task is due to the presence of the following shortcomings of the existing biometric models of users' keyboard handwriting: when the number of users in the system increases, the complexity of their classification increases; the complexity of formalizing the uniqueness of users; the implementation of determining the fact of replacing an already authorized user is too complex; the feature space of the existing biometric models is insignificant due to the limitation of the set of properties of the modern keyboard, which negatively affects the accuracy of the authentication procedure. A biometric model of keyboard handwriting was chosen, the feature of which is the formalization of the uniqueness of the user of the information system on the basis of the detection of the regularities of the keyboard handwriting by fuzzy rules. An improved method of regularization of the feature space of the biometric model of keyboard handwriting of users of military information systems is proposed. The essence of the proposed method, which distinguishes it from the existing ones, is that the increase in the set of features of the biometric model is achieved by adding to it discovered hidden facts from the set of own features based on factor analysis with the highest rate of their variability. The application of the proposed method allows solving the problem of non-linear separability of the n -dimensional feature space of the biometric model of keyboard handwriting, which in turn allows to increase the accuracy and reliability of the authentication procedure by control systems and access demarcation of military information systems.

Keywords: regularization, biometric model, behavioral biometrics, keyboard handwriting, information systems, authentication, unauthorized access, factor analysis.

Актуальність та постановка завдання в загальному вигляді. Існуючі умови технологічної ескалації у кіберпросторі залишають не вирішеним завдання ефективного забезпечення конфіденційності, доступності та цілісності даних інформаційних систем (ІС) критичної інфраструктури, зокрема ІС військового призначення [1–4].

Одним із основних напрямків вирішення даного завдання є організація виявлення фактів несанкціонованого доступу до інформаційних ресурсів (секретних даних, конфіденційної

інформації, оперативної обстановки військ у зоні бойових дій, місць дислокації засобів протиповітряної оборони, радіолокаційних станцій і т. д.), розголошення яких може призвести до непередбачуваних наслідків. Поряд з цим, існуючі алгоритми, способи, методи, методики автентифікації, як підгрунтя основного етапу процедури контролю доступу користувачів ІС не здатні повною мірою забезпечити ефективний кіберзахист інформаційних ІС, про що свідчать численні факти компрометації безпеки ІС та несанкціонованих втручань [2; 5].

Це обумовлює актуальність подальших наукових досліджень щодо підвищення ефективності процедури автентифікації користувачів ІС військового призначення.

Аналіз наукових публікацій показав ефективність використання процедури автентифікації користувачів ІС, в основу якої покладено аналіз їхньої поведінкової біометрії, зокрема клавіатурного почерку (КП), оскільки дозволяє здійснювати класифікацію (розпізнавання) користувачів за їх індивідуальними (унікальними) підсвідомими характеристиками (сенсорними і руховими навичками), які практично неможливо (повністю неможливо) фальсифікувати [2; 3]. Так, автентифікація користувачів відповідно до пред'явленого ідентифікатора здійснюється на основі аналізу значень показників машинного набору тексту (швидкість друку, ритм друку, сила натискання, тривалість натискання, час між натисканням клавіш) під час виконання процесу, що підлягає дослідженню (введення пароля, контрольного тексту).

Поряд з цим, існуючі системи біометричної автентифікації за КП, часто характеризуються досить низькою достовірністю (точністю) автентифікації осіб [6]. У переважній більшості наукових праць [3; 6–14], які присвячені вивченню даного питання, на відміну від завдання вибору науково-методичного апарату для подальшої побудови класифікатора, недостатньо уваги приділяється формалізації унікальності (індивідуальних підсвідомих характеристик) користувачів ІС, що зі свого боку негативно впливає на адекватність їхньої біометричної моделі (профілю) КП. Як правило, синтез біометричної моделі (профілю) КП здійснюється на основі наступних підходів:

шляхом визначення персонального ознакового простору для кожного користувача ІС;

шляхом зведення множини досліджуваних ознак КП користувачів ІС у спільний ознаковий простір.

Перший підхід передбачає формалізацію індивідуальних підсвідомих характеристик користувача досить вузько, оскільки характеризує конкретну послідовність натискання клавіш користувачем (контрольного тексту, пароля), що не відповідає реальній криміналістичній експертизі рукописного почерку людини [15]. Другий підхід, навпаки, дозволяє синтезувати біометричні моделі користувачів на основі їх спільного ознакового простору із забезпеченням можливості аналізу натискання довільних послідовностей клавіш клавіатури користувачем, що у значно більшій мірі відповідає реальній криміналістичній експертизі рукописного почерку користувача [15]. До того ж, такий підхід дозволяє виявляти підміну користувача ІС за іншим наявним ідентифікатором у системі.

Підходи до синтезу біометричних моделей КП шляхом визначення персонального ознакового простору для кожного користувача представлено значною кількістю наукових досліджень [3; 6–14], у переважній більшості яких [6–14] формування ознакового простору обмежується лише статистичними даними роботи з клавіатурою, що породжує низку недоліків:

при збільшенні числа користувачів у системі зростає складність їх класифікації;

складна реалізація формалізації унікальності користувачів ІС;

надто складна реалізація визначення факту підміни уже авторизованого користувача;

ознаковий простір існуючих біометричних моделей (профілів) є досить незначним в силу обмеження множини властивостей сучасної клавіатури, що негативно впливає на показник точності (достовірності) процедури автентифікації.

У роботі [3] запропоновано підхід до синтезу біометричної моделі КП користувача ІС, який ґрунтується на використанні математичного апарату теорії нечіткої логіки. Так, додавання нових ознак ґрунтується на основі інженерії поведінкових закономірностей роботи

користувача з клавіатурою у вигляді нечітких правил на множині досліджуваних ознак, що дозволяє певною мірою описати його унікальність. Даний підхід передбачає формування спільного ознакового простору для всіх біометричних моделей користувачів, однак не вирішує питання його нелінійної сепарабельності.

Представлені недоліки розглянутих наукових досліджень обумовлюють доцільність регуляризації ознакового простору існуючих біометричних моделей користувачів ІС, що дозволить підвищити ефективність їх класифікації (розпізнавання) на етапі процедури автентифікації.

Регуляризація (англ. *regularization* – метод додавання деяких додаткових даних до умови з метою вирішення некоректно поставленого завдання або додаткових обмежень до умови з метою запобігти перенаванченню [16]). У контексті поставленого наукового завдання – метод додавання певних додаткових ознак до ознакового простору біометричної моделі КП користувача ІС [3].

У зв'язку з цим, виникає актуальне наукове завдання регуляризації ознакового простору біометричної моделі КП користувачів ІС військового призначення.

Метою статті є розробка методу регуляризації ознакового простору біометричної моделі КП користувачів ІС військового призначення.

Біометрична модель КП користувачів ІС. У статті розглядається біометрична модель КП користувача ІС, запропонована у [3], синтез якої передбачає формування єдиного ознакового простору для всіх користувачів, що дозволяє виявляти підміну ідентифікованого користувача за іншим наявним ідентифікатором у системі. Застосування даної біометричної моделі КП дозволяє визначати притаманні конкретному користувачу підсвідомі унікальні поведінкові риси, присутні у різних психоемоційних станах, що у свою чергу дозволяє позбутися множини опису станів кожного облікового запису та зменшити кількість хибних спрацювань системою контролю та розмежування доступу у процесі автентифікації особи в умовах деякої нечіткості управляючої інформації.

Так, у відповідності до [3] із множини статистичних параметрів КП:

Δt_k^r – тривалість утримання (*retention*) клавіш k клавіатури;

Δt_{kl}^p – час натискання (*pause*) між клавішами (k та l);

pow – сила натискання клавіш (для сенсорної клавіатури);

sq – площа, що займається під час натискання клавіші (для сенсорної клавіатури),

першим кроком є синтез початкового спільного для всієї множини користувачів $u_i \in U$ n -вимірною ознакового простору КП S_{start} , де:

p – загальний відсоток попадання часового інтервалу утримання кожною клавішею послідовності у еталонні діапазони значень на основі відстані Махаланобіса [17; 18];

d – динаміка друку – час між натисканням клавіш і часом їх утримання;

v – швидкість друку – кількість натиснутих клавіш розділена на час друку;

i – ідентифікатор користувача ІС.

Другим кроком формування біометричної моделі є додавання до початкового ознакового простору S_{start} підмножини нових ознак S_{new} , створених на основі знайдених поведінкових закономірностей зі статистичного набору даних роботи користувача ІС u_i за клавіатурою. Реалізація способу інженерії закономірностей передбачає опис послідовності часових інтервалів утримання клавіш Δt_k^r клавіатури відповідною їй функцією $f(\Delta t_1^r, \Delta t_2^r, \dots, \Delta t_n^r)$. Для прикладу, на рисунку 1 зображено графік функції поведінкової закономірності КП першого автора при введенні контрольного тексту (2 спроби) із 12 символів (Δt_k^r у мілісекундах). По горизонтальній осі представлено загальний час введення контрольної фрази у мілісекундах, по вертикальній – Δt_k^r даної послідовності у мілісекундах.

Як видно з рисунку 1, наявна закономірність тривалості утримання клавіш Δt_k^r виражена наближеними формами кривих, які описують її у 12 точках із відповідними значеннями, що є справедливим і для значної кількості таких спроб.

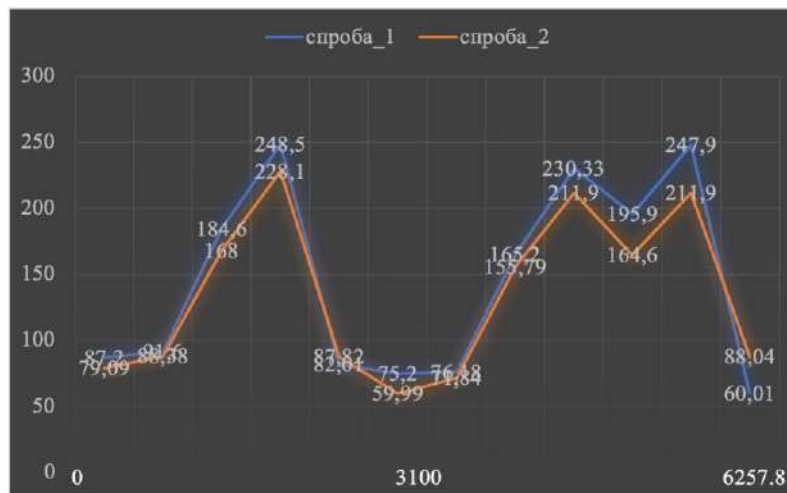


Рис. 1. Графік функції поведінкової закономірності КІ першого автора при введенні контрольної фрази (Δt_{1-12}^r)

Відповідно до [3] необхідно визначити додаткові ознаки КІ користувача на основі знайденої закономірності з метою формування остаточного ознакового простору біометричної моделі КІ M_{u_i} у поєднанні із початковим S_{start} . Так, представлена послідовність декомпозується на часові вікна t_i^w , які формують додаткову підмножину ознак КІ S_{new} (рис. 2). Кількість часових вікон – 10, що у середньому відповідає інтервалу 500 мілісекунд. Час введення власних паролів різними користувачами під час проведення експериментів не перевищував 10 секунд.

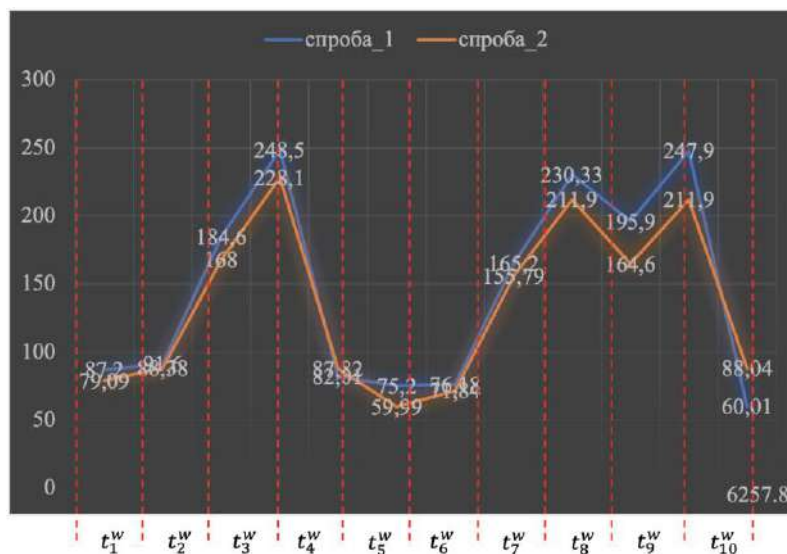


Рис. 2. Графічне представлення декомпозиції досліджуваної послідовності натискання клавіш користувачем u_i на часові вікна t_{1-10}^w .

Так, у кожному часовому вікні t_i^w присутні частини інтервалів (кусково-лінійні функції) виявленої закономірності. Вибір даного способу інженерії ознак зумовлено необхідністю інтерпретації послідовності утримання клавіш Δt_k^r клавіатури під час відтворення процесу, що підлягає обробці однаковою їх кількістю для різних користувачів u_i .

Заключним кроком синтезу біометричної моделі користувача є апроксимація функції $f(\Delta t_1^r, \Delta t_2^r, \dots, \Delta t_n^r)$, що описує швидкість зміни виявленої закономірності у вигляді кривої (геометричний сенс диференціювання функцій) нечіткими правилами у часових вікнах t_{1-10}^w (терм-множини: зростання (дуже високе, високе, середнє, мале, дуже мале); спадання (дуже високе, високе, середнє, мале, дуже мале)). При чому, визначення значення похідної

для будь-якої точки на кривій, що описує закономірність Δt_k^r ґрунтується на визначенні кута між горизонтальною віссю та дотичною до обраної точки.

Аналітичний вираз описаної біометричної моделі КП засобами теорії нечіткої логіки представлено аналітичним виразом (1) [3]:

$$M_{u_i} = (S_{start} = \{p, d, v\} \cup S_{new} = \{t_1^w, \dots, t_{10}^w\}) \rightarrow \{p, d, v, t_1^w, \dots, t_{10}^w\}. \quad (1)$$

Очевидно, що найбільш впливовим негативним чинником застосування даної моделі M_{u_i} для автентифікації користувачів ІС системами контролю і розмежування доступу є фактор нелінійної сепарабельності її ознакового простору КП, походження якого зумовлено залежністю складної нелінійної функції від кількох змінних, розподіл значення яких досить складно (неможливо) знайти. Іншими словами, немає таких функцій, які можуть розділити ці змінні на незалежні групи лінійно. На рисунку 3 представлено завдання класифікації векторів (червоні, сині), множини яких не можуть бути лінійно розділені на площині (рис. 3, а – випадок двох множин, рис. 3, б – випадок двох класів (кластерів)) [19].

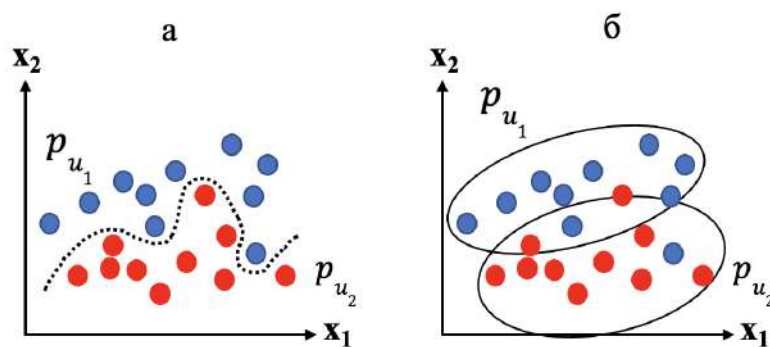


Рис. 3. Графічне представлення випадку нелінійної сепарабельності множин (кластерів) векторів: а – випадок двох множин; б – випадок двох класів (кластерів)

Рішення такого роду завдань існуючими підходами: методи оптимізації (градієнтний спуск, оптимізація на основі еволюційних алгоритмів); методи поділу змінних (лінійна регресія, аналіз головних компонент); методи на основі штучних нейронних мереж; ядрові методи, полягає у відображенні цільових вхідних даних на збільшеному просторі ознак з метою пошуку такої гіперплощини, яка дозволить розділити їх лінійно. На рис. 4, а зображено завдання та ідеальний результат його рішення (рис. 4, б) на збільшеному просторі ознак.

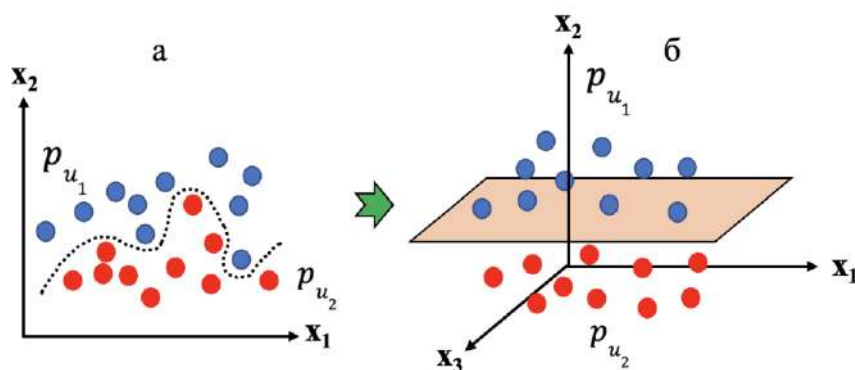


Рис. 4. Графічне представлення результату рішення завдання нелінійної сепарабельності у збільшеному просторі ознак: а – завдання; б – ідеальний результат його рішення

Застосування вищезазначених методів дозволяє вирішити завдання нелінійної сепарабельності ознакового простору лише частково, оскільки піднімає, як питання оптимального відбору інформативних ознак у процесі зменшення його розмірності, так і трансформації досліджуваних векторів з метою вирішення завдання їх класифікації на новому

просторі ознак. Даний факт спричиняє необхідність пошуку такого підходу до регуляризації ознакового простору біометричної моделі, який дозволить вирішити завдання нелінійної сепарабельності у цілому.

Метод регуляризації ознакового простору біометричної моделі КП. На основі викладеного пропонується удосконалення методу регуляризації біометричної моделі КП користувачів ІС військового призначення на основі факторного аналізу. Суть запропонованого методу, що відрізняє його від існуючих [3; 6–14] полягає у тому, що збільшення множини ознак біометричної моделі досягається шляхом додавання до неї виявлених нових прихованих фактів із множини власних ознак на основі факторного аналізу із найбільшим показником їх мінливості. Так, у результаті визначення структури взаємозв'язків досліджуваних змінних у вигляді ознак їх підмножина не заміщує досліджуваний ознаковий простір моделі, а використовуються для збільшення його вимірів.

Вибір факторного аналізу для вирішення завдання нелінійної сепарабельності ознакового простору зумовлено забезпеченням можливості формалізації взаємозв'язків змінних без поділу на результативні та факторні ознаки, наприклад, на відміну від регресійного аналізу, а також дисперсійного та дискримінантного у аспекті відсутності необхідності визначення заздалегідь управляючих даних.

Методом факторного аналізу обрано метод головних компонент [20] (англ. *Principal component analysis*, PCA) – метод факторного аналізу, який використовує ортогональне перетворення множини спостережень з пов'язаними змінними (сутностями, кожна з яких набуває різних числових значень) у множину змінних без лінійної кореляції, які називаються головними компонентами. Вибір методу PCA обумовлено тим фактом, що в аналізі головних компонент передбачається використання всієї мінливості змінних, тоді як, наприклад, в аналізі головних факторів використовується тільки мінливість конкретної змінної. До того ж, на практиці, PCA демонструє найкращі результати, як метод зменшення розмірності даних, тоді як інші методи краще застосовувати для визначення структури даних.

Формальна постановка завдання [21; 22]:

Дано центрований набір векторів даних $x_i \in R^n$ ($i = 1, \dots, m$) (середнє арифметичне значення дорівнює нулю). Вибіркова дисперсія даних вздовж напрямку, заданого нормованим вектором a_k , це (2):

$$S_m^2 [(X, a_k)] = \frac{1}{m} \sum_{i=1}^m (a_k, x_i)^2 = \frac{1}{m} \sum_{i=1}^m \left(\sum_{j=1}^n x_{ij} a_{kj} \right)^2. \quad (2)$$

Оскільки дані центровані, вибіркова дисперсія збігається із середнім квадратом відхилення від нуля. Рішення задачі найкращої апроксимації дає ту саму множину головних компонентів $\{a_j\}$, що і пошук ортогональних проєкцій з найбільшою дисперсією у зв'язку з тим, що перший доданок не залежить від a_k , а також $\|x_i - a_k(a_k, x_i)\|^2 = \|x_i\|^2 - (a_k, x_i)^2$.

Вихідні дані:

M_{u_i} – біометрична модель КП користувачів u_i ІС;

u_i – користувач ІС із множини усіх користувачів U ІС;

$\{p, d, v, t_1^w, \dots, t_{10}^w\}$ – 13-ознаковий простір M_{u_i} , на основі аналізу значень якого виконується завдання класифікації (розпізнавання) $u_i \in U$;

набір статистичних даних, на основі яких здійснювався синтез моделі M_{u_i} .

Розглядається процес регуляризації ознакового простору M_{u_i} з метою вирішення завдання його нелінійної сепарабельності.

Обмеження та допущення:

ознаковий простір M_{u_i} є спільним для усієї множини користувачів U у системі;

біометрична модель M_{u_i} описує роботу користувачів $u_i \in U$ із звичайною (традиційною) клавіатурою;

у випадку застосування сенсорної клавіатури до початкового ознакового простору M_{u_i} необхідно додати ознаки: сили натискання клавіш та площі, що займається під час натискання клавіш.

Необхідно:

знайти таке ортогональне перетворення (проекцію) ознакового підпростору M_{u_i} в нову систему координат S_{sub} , для якого були б справедливі наступні умови:

вибіркова дисперсія даних уздовж першої координати максимальна (перша головна компонента);

вибіркова дисперсія даних уздовж другої координати максимальна за умови ортогональності першої координати (друга головна компонента);

вибіркова дисперсія даних уздовж значень k -ї координати максимальна за умови ортогональності першим $k - 1$ координатам.

Метод регуляризації ознакового простору біометричної моделі КП користувачів ІС представлено наступною функціональною схемою, яка передбачає реалізацію наступних етапів (рис. 5):

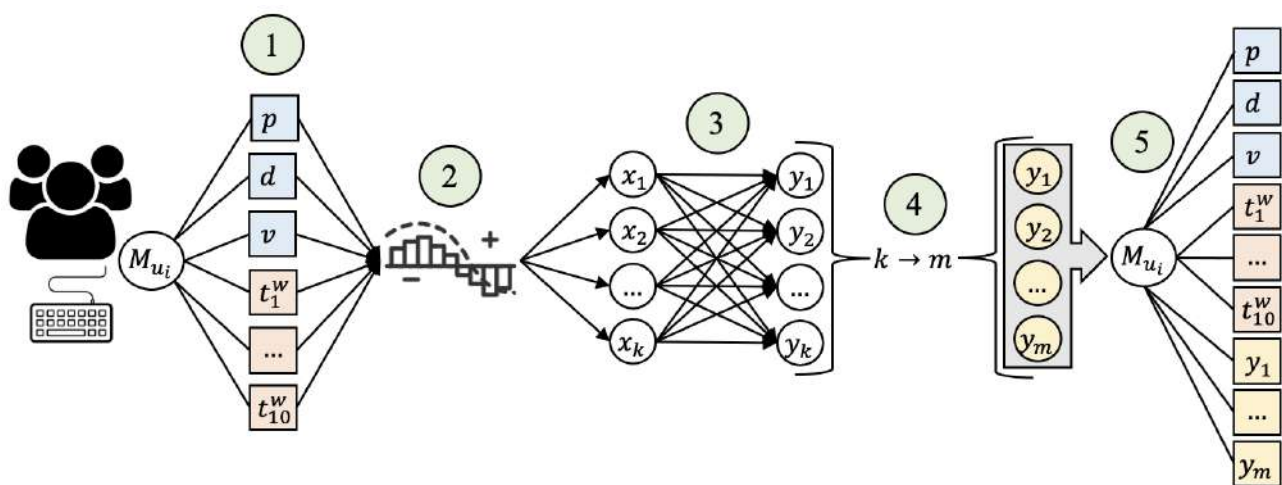


Рис. 5. Функціональна схема методу регуляризації ознакового простору біометричної моделі КП користувачів ІС

1. Отримання множини ознак біометричної моделі КП користувача ІС.

На даному етапі ініціалізується множина ознак біометричної моделі M_{u_i} потужністю 13, якою описуються користувачі ІС з метою ортогональної її проєкції на ознаковий простір зменшеної розмірності S_{sub} .

Для конструювання основних компонентів використовується лінійна модель для стандартизованих змінних [23] (3):

$$y_i = \sum_{j=1}^k a_{i,j} x_j, \quad (3)$$

де y_i – головна компонента за номером i ($i = 1, \dots, m$);

x_j – стандартизована вихідна змінна z_j ;

\bar{z}_j – середнє арифметичне ознаки ($j = 1, \dots, k$);

$a_{i,j}$ – коефіцієнт відображення значущості ознаки x_j у головну компоненту y_i .

2. Стандартизація значень вихідного ознакового простору.

На даному етапі здійснюється перехід від вихідного простору ознак $\{p, d, t_s, t_1^w, \dots, t_{10}^w\}$ до простору стандартизованих змінних x_1, \dots, x_k у відповідності до аналітичної моделі центрування (4):

$$\left(x_i = \frac{z_j - \bar{z}_j}{s_j} \right), (j = 1, \dots, k). \quad (4)$$

Середнє арифметичне стандартизованих змінних дорівнює нулю ($x_j = 0$), дисперсія та стандартне відхилення дорівнюють одиниці ($s_1^2 = s_j = 1$). Отже, стандартизовані змінні x_1, \dots, x_k мають однаковий рівень інформативності, а сумарний обсяг інформації, що в них міститься дорівнює (5):

$$k \left(\sum_{j=1}^k s_j^2 = k \right). \quad (5)$$

Однак, перед здійсненням стандартизації ознак необхідно реалізувати кодування значень векторів, оскільки біометрична модель M_{u_i} синтезована засобами теорії нечіткої логіки. Так, значення векторів конкретного користувача $u_i \in U$, що представляють собою терми нечіткої множини є категоріальними, а пошук прихованих взаємозв'язків у даних вимагають наявності числових значень.

3. Лінійне перетворення стандартизованого простору ознак.

На даному етапі здійснюється лінійне перетворення стандартизованого простору ознак з метою побудови нового ортогонального простору y_1, y_2, \dots, y_k у відповідності до (6):

$$y_i = \sum_{j=1}^k a_{i,j} x_j, (i = 1, \dots, k), \quad (6)$$

де y_i – нова змінна за номером i ($i = 1, \dots, k$);

x_j – стандартизована змінна за номером j ($j = 1, \dots, k$);

$a_{i,j}$ – коефіцієнти переходу від набору змінних x_1, x_2, \dots, x_k до y_1, y_2, \dots, y_k .

Лінійну модель для стандартизованих змінних на даному етапі доцільно представити у вигляді системи рівнянь (7):

$$\begin{cases} y_1 = a_{1,1}x_1 + a_{1,2}x_2 + \dots + a_{1,k}x_k \\ \dots \\ y_k = a_{k,1}x_1 + a_{k,2}x_2 + \dots + a_{k,k}x_k \end{cases} \quad (7)$$

Коефіцієнти $a_{i,j}$ розраховуються із врахуванням наступних умов:

значення дисперсії нових змінних y_i кількісно рівні власним значенням вихідної матриці кореляцій $s^2(y_i) = \lambda_i$. Сума власних значень матриці кореляцій дорівнює k , а отже, інформація, що міститься в множині стандартизованих змінних x_1, x_2, \dots, x_k , повністю зберігається у наборі нових змінних y_1, y_2, \dots, y_k ;

змінні y_1, y_2, \dots, y_k пронумеровані у порядку зменшення значень дисперсії $s^2(y_1) \geq s^2(y_2) \geq \dots \geq s^2(y_k)$.

Змінні y_1, y_2, \dots, y_k ортогональні, оскільки не корелюють одна з одною.

Таким чином, отримано новий простір ознак КП y_1, y_2, \dots, y_k , розмірність якого збігається з розмірністю вихідного простору. Новий простір ортогональний, змінні якого упорядковано за значенням зменшення їх дисперсії.

4. Визначення кількості головних компонент.

На даному етапі здійснюється зменшення ознакового простору біометричної моделі за допомогою "відсікання" певної кількості найменш інформативних змінних з максимальними номерами. Решта m змінних простору y_1, y_2, \dots, y_m ($m \ll k$) називаються головними компонентами. Так, в процесі визначення головних компонент система k рівнянь (7) зводиться до m рівнянь (8):

$$y_i = \sum_{j=1}^k a_{i,j} x_j, (i = 1, \dots, m). \quad (8)$$

Очевидно, що сума дисперсії головних компонент менша, ніж сума дисперсії вихідних змінних k , тому визначення кількості головних компонент доцільно здійснювати засобами критерію, заснованого на власних числах матриці кореляції. Суть даного підходу полягає в тому, щоб обмежити відбір головних компонент тими змінними y_i , яким відповідають власні значення $\lambda_i \geq 1$, тому що їх інформаційна значущість вища за інформаційну значущість відсічених змінних.

5. Синтез остаточного ознакового простору.

На даному етапі здійснюється отримання нового підпростору ознак КП S_{sub} шляхом інтерпретації максимальних за абсолютною величиною навантажень (коефіцієнтів лінійних перетворень $a_{i,j}$) із матриці навантажень у відповідності до кількості, значення якої визначено на попередньому етапі.

Матрицю навантажень на головні компоненти представлено таблицею 1.

Таблиця 1

Матриця навантажень на компоненти

| Ознака | y_1 | y_2 | ... | y_m | ... | y_k |
|--------|-----------|-----------|-----|-----------|-----|-----------|
| x_1 | $a_{1,1}$ | $a_{2,1}$ | ... | $a_{m,1}$ | ... | $a_{k,1}$ |
| x_2 | $a_{1,2}$ | $a_{2,2}$ | ... | $a_{m,2}$ | ... | $a_{k,2}$ |
| ... | ... | ... | ... | ... | ... | ... |
| x_k | $a_{1,k}$ | $a_{2,k}$ | ... | $a_{m,k}$ | ... | $a_{k,k}$ |

Таким чином, до ознакового простору біометричної моделі додається нечітка підмножина визначених головних компонент (9):

$$M_{u_i} = \{p, d, t_s, t_1^w, \dots, t_{10}^w\} \cup S_{sub} = \{y_1, \dots, y_m\} \rightarrow \{p, d, t_s, t_1^w, \dots, t_{10}^w, y_1, \dots, y_m\}. \quad (9)$$

Оцінка сепарабельності отриманого ознакового простору. Оскільки вирішення завдання регуляризації ознакового простору біометричної моделі КП M_{u_i} вирішує завдання її нелінійної сепарабельності, то оцінку отриманих результатів застосування методу доцільно здійснювати у рамках оцінки сепарабельності моделі до регуляризації її ознакового простору та після.

Оскільки, для n -вимірною ознакового простору множини векторів є лінійно-сепарабельні, якщо вони можуть бути відокремлені $(n-1)$ -вимірною гіперплощиною [24], а також із врахуванням того, що M_{u_i} побудована на основі теорії нечіткої логіки (нечітких множин), то оцінку її сепарабельності доцільно здійснювати на основі міри подібності Жаккара (коефіцієнту флористичної спільноти) [25; 26]. Суть підходу полягає у вимірюванні подібності множин на основі значення співвідношення міри спільної частини і міри їх об'єднання (10).

$$J(A, B) = 1 - J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}. \quad (10)$$

Таким чином, міра подібності Жаккара дозволить отримати значення зміни ознакового простору до регуляризації та після (від 0 до 1), де 0 – множини не мають спільних елементів, 1 – множини ідентичні.

Висновки. У статті вирішується актуальне наукове завдання регуляризації ознакового простору біометричної моделі КП користувачів ІС військового призначення у рамках підвищення ефективності процедури автентифікації користувачів системами контролю і розмежування доступу. Обрано біометричну модель КП, особливістю якої є формалізація унікальності користувача ІС на основі виявлення властивих йому закономірностей КП нечіткими правилами.

Запропоновано удосконалений метод регуляризації ознакового простору біометричної моделі КП користувачів ІС військового призначення. Суть запропонованого методу, що

відрізняє його від існуючих, полягає у тому, що збільшення множини ознак біометричної моделі досягається шляхом додавання до неї виявлених прихованих фактів із множини власних ознак на основі факторного аналізу із найбільшим показником їх мінливості.

Застосування даного підходу дозволяє вирішити описані у статті недоліки аналізованих наукових досліджень:

визначати факт підміни уже авторизованого користувача за наявними у системі ідентифікаторами;

вирішити питання нелінійної сепарабельності спільного для усіх користувачів системи ознакового простору;

поєднати переваги імовірісно-статистичних методів та методів машинного навчання для побудови біометричної моделі КП в умовах прийнятної обчислювальної складності.

Перспективним напрямком подальших наукових досліджень є розробка методики автентифікації користувачів інтелектуальними системами контролю та розмежування доступу ІС військового призначення.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Про національну безпеку України: Закон України від 21.06.2018 р. № 2469-VIII: станом на 31 берез. 2023 р.
2. Фесьоха В. В. Аналіз існуючих рішень автентифікації користувачів інформаційних систем та мереж спеціального призначення / В. В. Фесьоха, Н. О. Фесьоха, О. Д. Доброштан // Збірник наукових праць ВІТІ. 2020. № 3. С. 129–136.
3. Фесьоха В. В., Фесьоха Н. О. Модель нечіткої автентифікації користувачів інформаційних систем органів військового управління на основі поведінкової біометрії // Захист інформації. 2021. Т. 23, № 2. С. 116–123.
4. Фесьоха В. В., Кисиленко Д. Ю., Турчак О. Р. Перспективи удосконалення існуючих рішень виявлення шкідливого програмного забезпечення в інформаційних системах військового призначення // Системи і технології зв'язку, інформатизації та кібербезпеки: актуальні питання і тенденції розвитку: матеріали II Міжнар. наук.-практ. конф., м. Київ, 1 груд. 2022 р. Київ, ВІТІ ім. Героїв Крут, 2022. С. 216.
5. Zero-day polymorphic cyberattacks detection using fuzzy inference system / I. Y. Subach et al. // Austrian Journal of Technical and Natural Sciences. 2020. Vol. 5, 6. P. 8–13.
6. Алексеев В. А. Сравнительный анализ перспективных технологий аутентификации пользователей ПК по клавиатурному почерку / В. А. Алексеев, Д. В. Маслий, Д. Ю. Горелов // Радиотехника: Всеукр. межвед. науч.-техн. сб. 2017. Вып. 189. С. 195–201.
7. Shklyar E., Vorobyev E., Savelyev M. Browser-based keystroke dynamics recognition. Saint Petersburg Electrotechnical University «LETI». No 5/2019. Pp. 58–63.
8. Young J. R. and Hammon R. W. Method and Apparatus for Verifying an Individual's Identity. Patent Number 4,805,222, U.S. Patent and Trademark Office, Washington, D.C., Feb., 1989.
9. Kim J., Kang P. Recurrent neural network-based user authentication for freely typed keystroke data // arXiv preprint arXiv: 1806.06190. 2018.
10. Continuous authentication by free-text keystroke based on CNN and RNN / X. Lu et al. // Computers & Security. 2020. Vol. 96. P. 101861. URL: <https://doi.org/10.1016/j.cose.2020.101861> (date of access: 02.03.2023).
11. Yevetskyi V., Horniichuk I. Analysis of stability of the user's keyboard handwriting characteristics in the biometric authentication systems // Collection "Information technology and security". 2018. Vol. 6, no. 2. P. 19–28. URL: <https://doi.org/10.20535/2411-1031.2018.6.2.153487> (date of access: 05.03.2023).
12. Krutovostov D., Khitsenko V. Password Authentication and Continuous Authentication by Keystroke Dynamics Using Mathematical Statistics // Voprosy kiberbezopasnosti. 2017. № 5 (24). P. 91–99.
13. Чалая Л. Модель идентификации пользователей по клавиатурному почерку // Штучний інтелект. 2004. Т. 4. С. 811–817.
14. Heuristic Methods for Reservoir Monthly Inflow Forecasting: A Case Study of Xinfengjiang Reservoir in Pearl River, China / С.-Т. Cheng et al // Water. 2015. Vol. 7, no. 12. P. 4477–4495.
15. Судова почеркознавча експертиза // Київський науково-дослідний інститут судових експертиз. URL: <https://kndise.gov.ua/pocherkoznavcha/> (дата звернення: 15.03.2023).

16. L. Rosasco, T. Poggio, A Regularization Tour of Machine Learning, MIT-9.520 Lectures Notes (book draft), 2015.
17. Character recognition using the mahalanobis distance // Taguchi's quality engineering handbook. Hoboken, NJ, USA. С. 1288–1292.
18. Krutohvastov D., Khitsenko V. Password Authentication and Continuous Authentication by Keystroke Dynamics Using Mathematical Statistics // Voprosy kiberbezopasnosti. 2017. № 5 (24). P. 91–99.
19. Фесьоха В., Фесьоха Н. Удосконалена процедура автентифікації користувачів інформаційних систем військового призначення на основі аналізу біометричного профілю клавіатурного почерку // InterConf: матеріали 8-ї Міжнар. науково-практ. конф. «Глоб. Та регіон. Аспекти сталого розвитку», м. Копенгаген, 26–28 берез. 2023 р. 2023. С. 505–506.
20. Nikitin V. V., Bobin D. V. Principal component analysis for weighted data in the procedure of multidimensional statistical forecasting // Statistics and economics. 2021. Т. 18, № 2. С. 4–11.
21. Халимов Г. Анализ методов гарсия и главных компонент биометрической идентификации личности // Системи обробки інформації. 2013. № 5. С. 106–110.
22. Jolliffe I. T. Principal component analysis. 2nd ed. New York: Springer, 2002. 487 p.
23. Principal manifolds for data visualization and dimension reduction / ed. By A. N. Gorban et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008.
24. Сайфуллин Р., Александров С. Этапы реализации метода главных компонент при обработке сигналов аналитических приборов // Вестник самарского государственного технического университета. Серия «технические науки». 2018. Т. 26, № 2. С. 60–66.
25. Gruzling N. Linear separability of the vertices of an n-dimensional hypercube. Northern British Columbia: The University of Northern British Columbia, 2006. 88 p.
26. Jaccard P. Distribution de la flore alpine dans le Bassin des Dranses et dans quelques regions voisines // Bull. Soc. Vaudoise sci. Natur. 1901. V. 37, Bd. 140. S. 241–272.