

УДК 378.147.018.43:004.85:004.056

doi: 10.32620/reks.2021.1.06

А. С. ДОВБИШ, І. В. ШЕЛЕХОВ, Ю. О. ХІБОВСЬКА, О. В. МАТЯШ

Сумський державний університет, Україна

ІНФОРМАЦІЙНО-АНАЛІТИЧНА СИСТЕМА ОЦІНЮВАННЯ ВІДПОВІДНОСТІ СУЧАСНИМ ВИМОГАМ НАВЧАЛЬНОГО КОНТЕНТУ СПЕЦІАЛЬНОСТІ КІБЕРБЕЗПЕКА

Розв'язана актуальна задача підвищення функціональної ефективності машинного навчання інформаційно-аналітичної системи (ІАС) оцінки відповідності сучасним вимогам контенту навчальних дисциплін бакалаврського рівня спеціальності «Кібербезпека». Розроблено метод інформаційно-екстремального машинного навчання ІАС з метою адаптації навчального контенту випускової кафедри до вимог ринку праці, що дозволяє при функціонуванні системи в режимі моніторингу оперативно корегувати контент з навчальних дисциплін випускової кафедри. Ідея методу полягає у максимізації інформаційної спроможності ІАС в процесі машинного навчання, що дозволяє в режимі моніторингу отримати максимальну повну ймовірність прийняття правильних класифікаційних рішень. За результатами інформаційно-екстремального машинного навчання в рамках геометричного підходу побудовано вирішальні правила, практично інваріантні до багато вимірності простору ознак розпізнавання. Як критерій оптимізації параметрів машинного навчання використовується модифікація інформаційної міри Кульбака, яка є функціоналом точнісних характеристик класифікаційних рішень. Як параметри оптимізації розглядалися геометричні параметри гіперсферичних контейнерів класів розпізнавання, які в процесі машинного навчання відновлювалися в радіальному базисі бінарного простору ознак Хеммінга. При цьому вхідна навчальна матриця трансформувалася в робочу бінарну навчальну матрицю, яка змінювалася в процесі машинного навчання шляхом допустимих перетворень з метою адаптації вхідного математичного опису системи до максимальної достовірності класифікаційних рішень. Запропоновано категорійну модель функціонування ІАС, на основі якої розроблено алгоритм інформаційно-екстремального машинного навчання системи з автоматичним визначенням базового класу розпізнавання. За результатами опитування фахівців з кібербезпеки сформовано вхідну структуровану навчальну матрицю, а за результатами фізичного моделювання підтверджено працездатність запропонованого методу інформаційно-екстремального машинного навчання ІАС.

Ключові слова: інформаційно-екстремальна інтелектуальна технологія; машинне навчання; інформаційний критерій; оптимізація; навчальний контент; моніторинг; кібербезпека.

Вступ

Згідно з Європейськими освітніми стандартами якість освіти визначається користю, яку отримають як роботодавці, так і випускники вищого закладу освіти. Тому встановлення сталого зв'язку між випусковою кафедрою закладу вищої освіти та роботодавцями є необхідною умовою підвищення якості навчального процесу. Важливою складовою якості навчального процесу є навчальний контент випускової кафедри вищого закладу освіти. При цьому основним шляхом оцінювання навчального контенту є анкетування роботодавців. На практиці оброблення результатів опитування респондентів здійснюється із застосуванням методів багатовимірного статистичного аналізу, що потребує великих обсягів вхідних даних і відповідно значних матеріальних та часових витрат.

Одним із перспективних напрямів організації сталого моніторингу якості освіти є створення інформаційно-аналітичної системи (ІАС) для адаптації навчального контенту випускової кафедри до вимог ринку праці. Тому створення ІАС, здатної автоматично за результатами опитування респондентів формувати вхідний математичний опис і в режимі моніторингу оцінювати відповідність навчального контенту випускової кафедри сучасним вимогам є актуальною задачею.

В статті розглядається метод інформаційно-екстремального машинного навчання ІАС адаптації навчального контенту випускової кафедри до вимог ринку.

Огляд літератури

У праці [1] наголошується про важливість оптимізації навчального контенту як важливий

напряму підвищення якості навчального процесу, а одним із проблемних і складних завдань підвищення якості надання освітніх послуг є створення системи керування освітніми процесами з урахуванням вимог роботодавців до сучасних фахівців. Перспективним шляхом вирішення цієї проблеми є впровадження дуальної системи професійної освіти [2]. При цьому важливим ланцюгом автоматизованої системи керування дуальним навчанням є створення оперативного зворотного зв'язку між роботодавцями та випусковою кафедрою вищого закладу освіти. У праці [3] розглядалося питання автоматизації формування вхідного математичного опису інформаційно-аналітичної системи (ІАС) оцінювання відповідності навчального контенту сучасним вимогам. Аналіз сучасних підходів і тенденцій розвитку ІАС адаптації навчального контенту випускової кафедри до сучасних вимог шляхом моніторингу показав, що для підвищення достовірності рішень, що приймаються, необхідно використання інтелектуальних технологій аналізу даних [4, 5]. При цьому застосування існуючих методів інформаційного синтезу інтелектуальних ІАС, які базуються на відомих методах технології Data Minin [6, 7], включаючи штучні нейронні мережі [8, 9], ускладнюється на практиці через такі основні причини:

- довільні початкові умови процесу оцінювання;
- перетин у просторі ознак класів розпізнавання, які характеризують відповідні рівні якості навчального контенту;
- багато вимірність словника ознак і класів розпізнавання;
- вплив на машинне оцінювання неконтрольованих збурюючих факторів.

У працях [10, 11] наведено приклади створення експертних систем на основі нечіткої логіки та експертно-статистичних алгоритмів. Основні недоліки застосування таких систем для оцінювання якості освіти полягають в тому, що вони є негнучкими при перенавчанні і не забезпечують в режимі моніторингу зворотний зв'язок між випусковою кафедрою, роботодавцями та студентами різних форм навчання. Крім того, експертні системи не здатні автоматично формувати вхідний математичний опис ІАС і базу знань.

Таким чином, науково-методологічні питання створення ІАС оцінки якості навчального контенту все ще залишаються недостатньо дослідженими через ускладнення науково-методологічного характеру. Тому на практиці для переважної більшості навчальних закладів основним методом адаптації навчального контенту випускових кафедр до вимог ринку праці все ще залишається анкетування, аналіз

результатів якого потребує значних матеріальних і часових витрат.

Як перспективний шлях створення інтелектуальних систем аналізу результатів опитування респондентів є використання ідей і методів створеної в Сумському державному університеті так званої інформаційно-екстремальної інтелектуальної технології (ІЕІ-технології) [12, 13]. Запропоновані в рамках цієї технології методи інтелектуального аналізу даних базуються на принципі максимізації інформаційної спроможності системи в процесі її машинного навчання. При цьому накопичений досвід застосування методів інформаційно-екстремального машинного навчання інтелектуальних систем різного призначення показав, що для забезпечення високої функціональної ефективності машинного навчання необхідно враховувати специфічність вхідних даних. Для ІАС оцінки якості навчального контенту такою специфікою є наявність впорядкованого алфавіту класів розпізнавання, які характеризують різні рівні якості навчального контенту. Цей факт ставить перед розробником інформаційного та алгоритмічного забезпечення ІАС задачу вибору так званого базового класу розпізнавання, відносно якого встановлюється система контрольних допусків на ознаки розпізнавання.

Метою статті є підвищення функціональної ефективності ІАС адаптації навчального контенту випускової кафедри на прикладі освітньо-професійної програми спеціальності «Кібербезпека» першого (бакалаврського) рівня шляхом машинного навчання та розпізнавання образів.

Формалізована постановка задачі

Розглянемо в рамках ІЕІ-технології формалізовану постановку задачі інформаційного синтезу здатної навчатися ІАС адаптації навчального контенту випускової кафедри до вимог ринку праці. Нехай дано алфавіт $\{X_m^o \mid m = \overline{1, M}\}$ класів розпізнавання, які характеризують відповідний рівень якості навчального контенту спеціальності випускової кафедри. За оцінками респондентів, які виставляються по стобальній шкалі, сформовано навчальну матрицю типу «об'єкт властивість» $\|y_{m,i}^{(j)}\|, i = \overline{1, N}, j = \overline{1, n}$, де N – кількість ознак розпізнавання; n – кількість структурованих векторів ознак (далі в тексті просто реалізації) класів розпізнавання відповідно. При цьому рядок матриці $\{y_{m,i}^{(j)} \mid i = \overline{1, N}\}$ визначає j -ту реалізацію, а стовпчик $\{y_{m,i}^{(j)} \mid j = \overline{1, n}\}$ – навчальну випадкову вибірку значень i -тої ознаки.

Відомо, що концепція ІЕІ-технології полягає в перетворенні вхідної навчальної матриці Y в робочу бінарну матрицю X , яка шляхом допустимих перетворень в процесі машинного навчання адаптується до максимальної повної ймовірності прийняття правильних класифікаційних рішень. Тому для бінарного простору Хеммінга задано структурований вектор параметрів функціонування, які впливають на функціональну ефективність машинного навчання ІАС. В подальшому такі параметри функціонування будемо називати параметрами машинного навчання. Вектор параметрів машинного навчання ІАС розпізнавання реалізації класу X_m^o представимо у вигляді структури

$$g_m = \langle x_m, d_m, \delta_k \rangle, \quad (1)$$

де x_m – усереднений структурований вектор ознак класу розпізнавання X_m^o ;

d_m – радіус гіперсферичного контейнера класу розпізнавання X_m^o , який відновлюється в радіальному базисі простору ознак розпізнавання;

δ – параметр поля контрольних допусків на ознаки розпізнавання.

В методах інформаційно-екстремального машинного навчання оптимізація системи контрольних допусків є обов'язковою процедурою. При цьому параметр δ дорівнює половині симетричного поля контрольних допусків на ознаки розпізнавання як це показано на рис. 1.

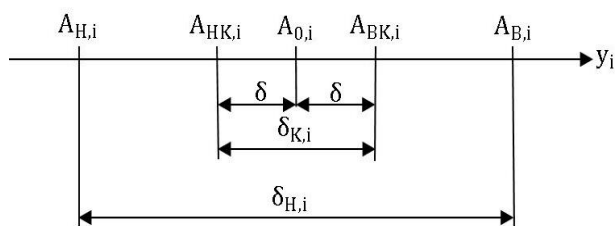


Рис. 1. Поля допусків на ознаку розпізнавання

На рис. 1 прийнято такі позначення: $A_{0,i}$ – номінальне значення ознаки y_i ; $A_{H,i}$ – нижній нормований (експлуатаційний) допуск; $A_{B,i}$ – верхній нормований допуск; $A_{HK,i}$ – нижній контрольний допуск; $A_{BK,i}$ – верхній контрольний допуск; $\delta_{k,i}$ – поле контрольних допусків; $\delta_{H,i}$ – поле нормованих допусків.

Двобічне симетричне поле контрольних допусків через параметр δ визначається за формулою

$$\delta_{k,i} = 2\delta \frac{A_{B,i} - A_{H,i}}{a},$$

де a – кількість градацій контрольного поля допусків, яка для всіх ознак розпізнавання є однаковою.

Область значень радіуса контейнера класу розпізнавання X_m^o задається нерівністю

$$d_m < d(x_m \oplus x_c),$$

де $d(x_m \oplus x_c)$ – міжцентрова відстань найближчих класів розпізнавання X_m^o і X_c^o , яка визначається як кодова відстань між відповідними векторами ознак x_m і x_c .

Обмеження на параметр δ задається нерівністю

$$0 < \delta < \delta_H / 2,$$

де δ_H – нормоване поле допусків, яке задає область значень контрольного поля допусків.

Необхідно в процесі машинного навчання ІАС:

1. Визначити оптимальні значення параметрів машинного навчання (1), які забезпечують максимум усередненого за алфавітом класів розпізнавання інформаційного критерію

$$\bar{E}^* = \frac{1}{M} \sum_{m=1}^M \max_{G_E \cap \{k\}} E_m^{(k)}, \quad (2)$$

де $E_m^{(k)}$ – інформаційний критерій оптимізації параметрів машинного навчання ІАС розпізнавати реалізації класу X_m^o , значення якого обчислюються на k -му кроці машинного навчання;

G_E – допустима область визначення функції інформаційного критерію оптимізації, яка далі буде називатися робочою областю;

$\{k\}$ – впорядкована множина кроків машинного навчання.

Для априорно класифікованого нечіткого розбиття $\tilde{\mathfrak{R}}^{[M]}$ побудувати шляхом допустимих перетворень в субпарацептуальному бінарному просторі ознак розпізнавання Хеммінга оптимальне, в інформаційному розумінні, чітке розбиття класів розпізнавання $\mathfrak{R}^{[M]}$, на основі якого сформулювати безпомилкові за навчальною матрицею вирішальні правила.

2. На етапі екзамену з метою перевірки функціональної ефективності машинного навчання прийняти рішення про належність реалізації образу,

що розпізнається, до одного із класів заданого алфавіту $\{X_m^o\}$.

Таким чином, задача інформаційного синтезу здатної навчатися ІАС зводиться до оптимізації в процесі інформаційно-екстремального машинного навчання параметрів функціонування (1) за інформаційним критерієм (2) і до прийняття в режимі екзамену класифікаційного рішення за побудованими на етапі навчання вирішальними правилами.

Категорійна модель

Категорійну модель інформаційно-екстремального машинного навчання ІАС представимо у вигляді спрямованого графу відображення операторами одна на одну відповідних множин, які застосовуються в процесі навчання.

Вхідний математичний опис здатної навчатися ІСА подамо у вигляді структури

$$I_{вх} = \langle G, T, \Omega, Z, Y, X; f_1, f_2 \rangle,$$

де G – простір вхідних факторів, які впливають на процес оцінювання навчального контенту;

T – множина моментів часу зняття інформації;

Ω – простір ознак розпізнавання;

Z – простір можливих станів ІАС;

Y – вхідна навчальна матриця;

X – робоча бінарна навчальна матриця, яка в процесі інформаційно-екстремального машинного навчання адаптується до максимальної повної ймовірності прийняття правильних класифікаційних рішень;

f_1 – оператор формування вхідної навчальної матриці Y із джерела інформації, яке задається декартовим добутком множин

$$G \times T \times \Omega \times Z;$$

f_2 – оператор перетворення вхідної навчальної матриці Y в робочу бінарну навчальну матрицю X .

Категорійну модель інформаційно-екстремального машинного навчання ІАС з оптимізацією контрольних допусків на ознаки розпізнавання показано на рис. 2.

На рис. 2 терм-множина E , яка складається із обчислених на кожному кроці машинного навчання значень інформаційного критерію, є загальною для всіх контурів оптимізації параметрів. Оператор $r: E \rightarrow \tilde{R}^{(M)}$ в процесі машинного навчання відновлює в радіальному базисі бінарного простору ознак контейнери класів розпізнавання, які утворюють розбиття $\tilde{R}^{(M)}$. Оператор θ відображає

розбиття $\tilde{R}^{(M)}$ на нечіткий розподіл апріорно класифікованих двійкових векторів ознак класів розпізнавання. Далі оператор $\psi: X \rightarrow I^{(S)}$, де $I^{(S)}$ – множина гіпотез, перевіряє основну статистичну гіпотезу $\gamma_1: x_m^{(i)} \in X_m^o$. Оператор γ визначає множину точнісних характеристик $\mathcal{Z}^{(Q)}$, де $Q = S^2$, а оператор ϕ обчислює множину значень E інформаційного критерію оптимізації, який є функціоналом від точнісних характеристик. Категорійна модель містить контур операторів оптимізації контрольних допусків на ознаки розпізнавання, який замикається через терм-множину D допустимих значень системи контрольних допусків. При цьому оператор δ_1 на кожному кроці машинного навчання змінює контрольне поле, а оператор δ_2 оцінює залежність ознак розпізнавання заданому контрольному полю допусків. Оператор u регламентує процес машинного навчання.

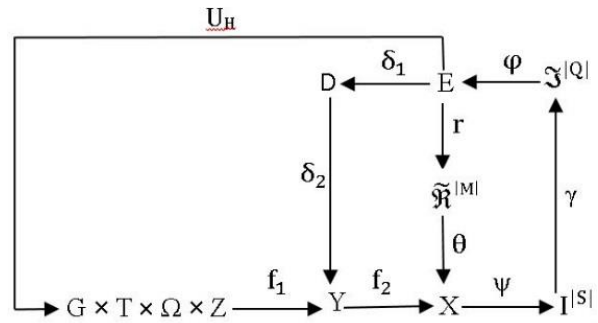


Рис. 2. Категорійна модель машинного навчання з оптимізацією контрольних допусків

Категорійна модель (див. рис. 2) відбиває притаманні людині перетворення інформації та інформаційні потоки, які мають місце при когнітивних процесах формування та прийняття класифікаційних рішень. Тому її можна розглядати як узагальнену структурну схему алгоритмів інформаційно-екстремального машинного навчання.

Згідно з принципом відкладених рішень О. Г. Івахненка виникає необхідність збільшення глибини інформаційно-екстремального машинного навчання до тих пір, поки не буде побудовано безпомилкові за навчальною матрицею вирішальні правила. Оскільки в методах ІЕІ-технології глибина машинного навчання визначається кількістю параметрів машинного навчання, що оптимізуються, то у цьому випадку категорійна модель, показана на рис. 2, буде мати відповідно додаткові контури оптимізації. При цьому згідно з принципом повної композиції як обов'язкова умова терм-множина E повинна бути загальною для всіх контурів оптимізації.

Машинне навчання з визначенням базового класу розпізнавання для заданого алфавіту

Ефективність функціонування системи в режимі машинного навчання залежить від заданої системи контрольних допусків на ознаки, яка формується відносно базового класу розпізнавання. На практиці за базовий прийнято вважати клас, що є найбільш бажаним для особи, яка приймає рішення про функціональну ефективність системи. Але в методах інформаційно-екстремального машинного навчання вибрати базовий клас доцільно з міркувань досягнення максимальної інформаційної спроможності системи, яка визначає максимальну повну ймовірність прийняття класифікаційних рішень. Тому розглянемо алгоритм інформаційно-екстремального машинного навчання ІАС, який визначає базовий клас розпізнавання за максимальним усередненим за алфавітом $\{X_m^0\}$ значень інформаційного критерію (2). Згідно з категорійною моделлю (рис. 2) алгоритм інформаційно-екстремального машинного навчання ІАС з оптимізацією контрольних допусків на ознаки розпізнавання представимо у вигляді двох циклічної ітераційної процедури пошуку глобального максимуму інформаційного критерію в робочій області визначення його функції:

$$\{\delta_{K,i}^* \mid i = \overline{1, N}\} = \arg \max_{G_\delta} \{ \max_{G_{E \cap (k)}} \bar{E}^{(k)} \}, \quad (3)$$

де $\bar{E}^{(k)} = \frac{1}{M} \sum_{m=1}^M E_m^{(k)}$ – усереднене значення інформаційного критерію, обчислене на k -му кроці машинного навчання;

G_δ – область допустимих значень контрольних допусків на ознаки розпізнавання.

Оскільки значення ознак розпізнавання мають однакову шкалу виміру, то доцільно реалізувати алгоритм машинного навчання з паралельною оптимізацією системи контрольних допусків, за якою на кожному кроці машинного навчання контрольні допуски змінюються для всіх ознак розпізнавання одночасно.

Вхідною інформацією для алгоритму машинного навчання є масив навчальної матриці $\{y_{m,i}^{(j)}\}$ і система полів нормованих допусків $\{\delta_{H,i}\}$ на ознаки розпізнавання, яка задає область значень відповідних контрольних допусків.

Розглянемо основні етапи реалізації алгоритму машинного навчання ІАС за процедурою (3) з паралельною оптимізацією контрольних допусків на ознаки розпізнавання:

1) ініціалізація лічильника класів розпізнавання: $m=m+1$;

2) виконується оптимізація системи контрольних допусків за процедурою (3);

3) пошук найбільшого усередненого значення інформаційного критерію (2) в межах робочої області;

4) якщо $m \leq M$, то виконується перехід до пункту 1, інакше – до пункту 5;

5) вибір базового класу шляхом пошуку найбільшого усередненого значення критерію (2) для класів розпізнавання заданого алфавіту:

$$m^* = \arg \max_{\{m\}} \bar{E}^* ;$$

6) визначаються оптимальні координати структурованого вектора (1):

$\{x_m^* \mid m = \overline{1, M}\}$ – оптимальні усередненні вектори ознак класів розпізнавання із заданого алфавіту;

$\{d_m^* \mid m = \overline{1, M}\}$ – оптимальні радіуси контейнерів класів розпізнавання;

δ^* – оптимальний параметр поля контрольних допусків на ознаки розпізнавання;

7) за параметром δ^* обчислюються нижні $\{A_{HK,i}^* \mid i = \overline{1, N}\}$ і верхні $\{A_{BK,i}^* \mid i = \overline{1, N}\}$ оптимальні контрольні допуски на ознаки розпізнавання:

$$A_{HK,i}^* = y_{m^*} - \delta^* \quad \text{і} \quad A_{BK,i}^* = y_{m^*} + \delta^* ;$$

8) ЗУПИН.

При наявності стобальної шкали оцінювання для нижніх і верхніх контрольних допусків, які обчислюються в процесі машинного навчання ІАС, встановлюються обмеження

$$A_{H,i} \geq 0 \quad \text{і} \quad A_{B,i} \leq 0 .$$

Як критерій оптимізації параметрів машинного навчання розглядалася модифікована інформаційна міра Шеннона:

$$E_m^{(k)} = 1 + \frac{1}{2} \left(\frac{K_{1,m}^{(k)}(d)}{K_{1,m}^{(k)}(d) + K_{3,m}^{(k)}(d)} \log_2 \frac{K_{1,m}^{(k)}(d)}{K_{1,m}^{(k)}(d) + K_{3,m}^{(k)}(d)} + \frac{K_{2,m}^{(k)}(d)}{K_{2,m}^{(k)}(d) + K_{4,m}^{(k)}(d)} \log_2 \frac{K_{2,m}^{(k)}(d)}{K_{2,m}^{(k)}(d) + K_{4,m}^{(k)}(d)} + \frac{K_{3,m}^{(k)}(d)}{K_{1,m}^{(k)}(d) + K_{3,m}^{(k)}(d)} \log_2 \frac{K_{3,m}^{(k)}(d)}{K_{1,m}^{(k)}(d) + K_{3,m}^{(k)}(d)} + \frac{K_{4,m}^{(k)}(d)}{K_{2,m}^{(k)}(d) + K_{4,m}^{(k)}(d)} \log_2 \frac{K_{4,m}^{(k)}(d)}{K_{2,m}^{(k)}(d) + K_{4,m}^{(k)}(d)} \right), \quad (4)$$

де $K_{1,m}^{(k)}(d)$ – кількість подій, які означають належність “своїх” реалізацій класу розпізнавання X_m^o ;

$K_{2,m}^{(k)}(d)$ – кількість подій, які означають неналежність “своїх” реалізацій класу розпізнавання X_m^o ;

$K_{3,m}^{(k)}(d)$ – кількість подій, які означають належність “чужих” реалізацій класу розпізнавання X_m^o ;

$K_{4,m}^{(k)}(d)$ – кількість подій, які означають неналежність “чужих” реалізацій класу розпізнавання X_m^o ;

d – дистанційна міра, яка визначає радіуси гіперсферичних контейнерів класів розпізнавання, побудованих в радіальному базисі бінарного простору Хеммінга;

За отриманими в процесі машинного навчання оптимальними геометричними параметрами контейнерів класів розпізнавання будуються вирішальні правила, які в продукційній формі представимо у вигляді

$$(\forall X_m^o \in \mathfrak{R}^M) \left(\text{if } [(\mu_m > 0) \& (\mu_m = \max_{\{m\}} \{\mu_m\})] \right. \\ \left. \text{then } x^{(j)} \in X_m^o \text{ else } x^{(j)} \notin X_m^o \right), \quad (5)$$

де $x^{(j)}$ – вектор, що розпізнається;

μ_m – функція належності вектора $x^{(j)}$ контейнеру класу розпізнавання X_m^o .

Таким чином, вектор $x^{(j)}$, що розпізнається, відноситься до класу розпізнавання X_m^o за умови, що функція належності μ_m є знакододатною і є максимальною серед функцій належності для інших класів розпізнавання.

У виразі (5) функція належності для гіперсферичного контейнера класу розпізнавання X_m^o визначається за формулою

$$\mu_m = 1 - \frac{d(x_m^* \oplus x^{(j)})}{d_m^*}, \quad (6)$$

де $d(x_m^* \oplus x^{(j)})$ – кодова відстань між вектором x_m^* і вектором $x^{(j)}$, що розпізнається.

Перевірка функціональної ефективності машинного навчання здійснюється за вирішальними правилами (5) при функціонуванні ІАС в режимі екзамену. Розглянемо основні етапи алгоритму

екзамену, за яким система функціонує і в режимі моніторингу:

1) ініціалізація лічильника класів розпізнавання: $m=m+1$;

2) обчислення кодової відстані Хеммінга $d(x_m^* \oplus x^{(j)})$;

3) обчислення функції належності (6);

4) якщо $m \leq M$, то виконується перехід до пункту 1, інакше – пункт 5;

5) визначається максимальне значення функції належності $\max_{\{m\}} \{\mu_m\}$;

6) за вирішальними правилами (5) приймається рішення про належність вектора ознак $x^{(j)}$ одному із класів заданого алфавіту;

7) ЗУПИН.

Таким чином, алгоритм інформаційно-екстремального машинного навчання ІАС полягає в оптимізації параметрів машинного навчання шляхом цілеспрямованого ітераційного пошуку глобального максимуму інформаційного критерію (4) в робочій (допустимій) області визначення його функції. Побудовані в процесі машинного навчання в рамках геометричного підходу вирішальні правила (5) через незначну обчислювальну трудомісткість характеризуються високою оперативністю і практично інваріантні до багато вимірності словника ознак розпізнавання.

Приклад реалізації алгоритму машинного навчання

Реалізація вище наведеного інформаційно-екстремального алгоритму машинного навчання ІАС здійснювався на прикладі оцінки відповідності вимогам ринку праці навчального контенту із спеціальності «Кібербезпека» бакалаврського рівня, яка викладається студентам Сумського державного університету. Загальні програмні компетентності освітньо-професійної програми спеціальності було сформовано в рамках проекту Європейського союзу «Tuning Educational Structures in Europe» та стандарту вищої освіти України для першого (бакалаврського) рівня спеціальності 125 «Кібербезпека». Кожний із респондентів – фахівців ІТ-компаній і установ, які займаються питаннями кібербезпеки, визначали суб'єктивно важливість тільки того чи іншого модуля 25 навчальних дисциплін, аде не оцінювали програмні результати, встановленні стандартом вищої освіти.

Вхідна навчальна матриця формувалася шляхом автоматичного зчитування оцінок змістовних модулів навчальних дисциплін, які виставлялися

респондентами за стобальною шкалою. Рівні якості навчального контенту характеризувалися чотирма класами розпізнавання, для яких на стобальній шкалі оцінок було попередньо визначено такі нечіткі інтервали: до класу X_1^o «відмінно» відносилися вектори ознак, які попадали в інтервал від 85 до 100 балів; до класу X_2^o «добре» – від 75 до 90 балів; до класу X_3^o «задовільно» – від 60 до 78 балів і до класу X_4^o «незадовільно» – менше 65 балів. Навчальна матриця для кожного класу складалася із 40 векторів ознак розпізнавання. Кожний вектор ознак розпізнавання був структурований за навчальними дисциплінами і відповідно за ознаками розпізнавання, кількість яких дорівнювала 147.

Згідно з освітньо-професійною програмою спеціальності «Кібербезпека» навчальні дисципліни було розбито на сім блоків:

1. Блок «Загальнонаукові дисципліни» включає дисципліни: вища математика, організація та обробка електронної інформації, обслуговування комп'ютерної техніки і програмування.

2. Блок «Фундаментальні дисципліни» включає дисципліни: вступ до спеціальності, спеціальні розділи математики, основи сучасних Інтернет-технологій, теоретичні аспекти захищених інформаційно-комунікаційних технологій, математичні методи моделювання та оптимізації процесів кібербезпеки.

3. Блок «Проектувальні дисципліни»: технології безпечного програмування, алгоритми захисту інформації, системи та засоби криптоаналізу, безпека веб-ресурсів, безпека Java-додатків.

4. Блок «Технологічні дисципліни»: фізичні основи кібербезпеки, технічні заходи забезпечення інформаційної безпеки, безпека комп'ютерних мереж, захищені інформаційні системи та бази даних.

5. Блок «Організаційні дисципліни»: теорія ризиків, система стандартів інформаційної та кібербезпеки, комплексні системи захисту інформації: проектування, впровадження, супровід, управління інцидентами безпеки.

6. Блок «Гуманітарні дисципліни»: іноземна мова, інтегрований курс "Основи академічного письма", інтегрований курс "Основи академічного письма".

7. Блок «Вибіркові дисципліни», який складався з дисциплін, в яких викладалися сучасні інтелектуальні методи аналізу даних в кібербезпеці.

Рівень селекції координат двійкових усереднених векторів ознак для всіх класів розпізнавання дорівнював 0,5.

Попередньо шляхом послідовного прогону алгоритму машинного навчання за процедурою (3) було визначено як базовий клас розпізнавання X_2^o , відносно якого в процесі машинного навчання будувалося поле контрольних допусків. На рис. 3 показано графік залежності усередненого за алфавітом класів розпізнавання інформаційного критерію (4) від параметра δ , отриманий при паралельній оптимізації контрольних допусків.

На рис. 3 і далі темною ділянкою на графіку позначено робочу область визначення функції критерію (4), в якій помилки першого та другого роду менше відповідно першої та другої достовірностей класифікаційних рішень. Аналіз рис. 3 показує, що оптимальне значення параметра поля контрольних допусків дорівнює $\delta = 14$ (у градаціях стобальної шкали оцінювання) при максимальному значенні усередненого нормованого критерію оптимізації $\bar{E}^* = 0,96$.

Для побудови вирішальних правил (5) необхідно знання геометричних параметрів контейнерів класів розпізнавання.

На рис. 4 показано скріншоти графіків залежності критерію (4) від радіусів контейнерів класів розпізнавання, отриманих при оптимальному значенні параметра поля контрольних допусків.

Аналіз рис. 4 показує, що оптимальні радіуси контейнерів класів розпізнавання відповідно дорівнюють $d_1^* = 4$ (тут і далі в кодових одиницях), $d_2^* = 15$, $d_3^* = 4$ і $d_4^* = 13$ при максимальних значеннях критерію (4): $E_1^* = E_2^* = E_3^* = 1$.

При максимальному значенні критерію $\bar{E}^* = 0,96$, як це показано на рис. 3, перша і друга усереднені достовірності дорівнювали відповідно $\bar{D}_1^* = 0,98$ і $\bar{D}_2^* = 0,96$. При цих точнісних характеристиках обчислена згідно з принципом Бернуллі при рівно ймовірних апріорних гіпотезах усереднена за алфавітом класів розпізнавання асимптотична (потенційна) повна ймовірність прийняття правильних класифікаційних рішень, отримана за результатами машинного навчання, дорівнює $P_t^* = 0,97$.

Аналіз результатів машинного навчання показує, що побудовані вирішальні правила характеризуються високою достовірністю класифікації векторів ознак розпізнавання, але не є безпомилковими за навчальною матрицею. Для підвищення функціональної ефективності машинного навчання необхідно збільшити його глибину шляхом оптимізації додаткових параметрів функціонування ІАС.

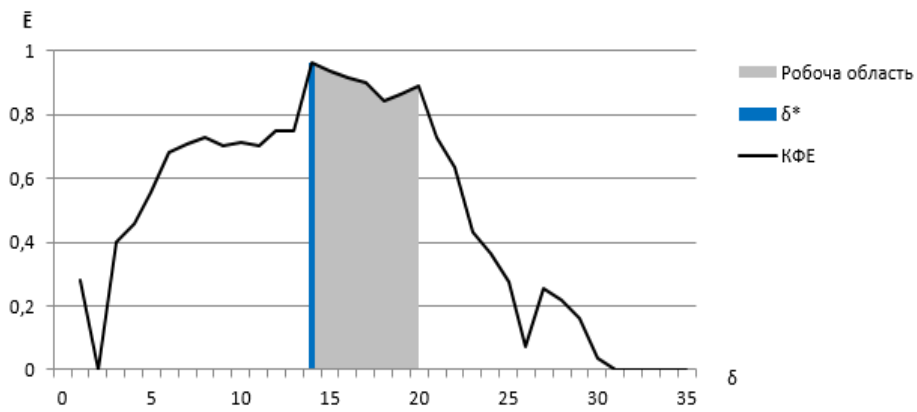


Рис. 3. Графік залежності інформаційного критерію від параметра поля контрольних допусків при базовому класі розпізнавання X_2^o

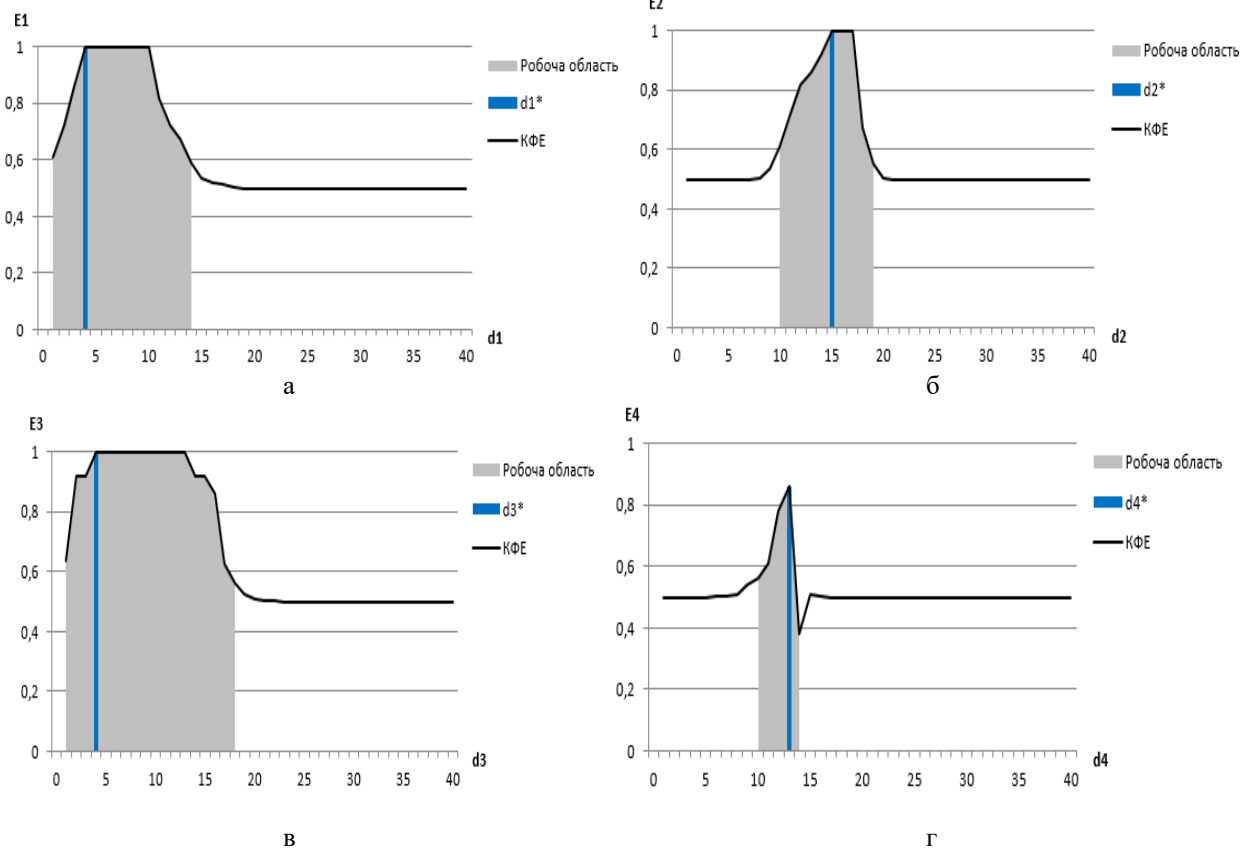


Рис. 4. Графіки залежності інформаційного критерію від радіусів контейнерів класів розпізнавання:
а – клас X_1^o ; б – клас X_2^o ; в – клас X_3^o ; г – клас X_4^o

Розглянутий метод інформаційно-екстремального машинного навчання дозволив побудувати ІАС першого покоління, задачею якої є загальна оцінка відповідності вимогам ринку праці навчального контенту із спеціальності «Кібербезпека» першого (бакалаврського) рівня. Використання такої системи в режимі моніторингу із щосеместровою періодичністю дозволяє здійснювати неперервну корекцію навчального контенту і розглядається як

обов'язкова інноваційна складова сучасного навчального процесу. Крім того, сама наявність на випусковій кафедрі вищого закладу освіти такої інтелектуальної ІАС в значній мірі виконує акредитаційні вимоги щодо взаємозв'язку випускової кафедри з роботодавцями.

Подальший розвиток синтезованої ІАС полягає в розширенні її функціональних можливостей, які дозволяють оцінювати відповідність сучасним вимогам як окремих блоків, так і самих навчальних

дисциплін за їх тематичними модулями, задекларованими у відповідних силабусах. При цьому із збільшенням потужності алфавіту класів розпізнавання виникає необхідність переходу від вище розглянутих лінійних алгоритмів інформаційно-екстремального машинного навчання до використання ієрархічних структур даних при функціонуванні ІАС в режимі факторного кластер-аналізу [11, 12].

Висновки

1. Розроблений метод машинного навчання ІАС згідно з Європейськими освітніми стандартами і національним стандартом вищої школи дозволяє випусковій кафедрі вищого закладу освіти шляхом машинного аналізу результатів опитування респондентів в режимі моніторингу оцінювати відповідність навчального контенту сучасним вимогам і є важливою складовою підвищення якості навчального процесу.

2. Запропоновано новий метод інформаційно-екстремального машинного навчання з автоматичним визначенням базового класу розпізнавання, відносно якого будується система контрольних допусків на ознаки розпізнавання, що дозволило побудувати високо достовірні вирішальні правила.

3. Реалізація запропонованого методу інформаційно-екстремального синтезу ІАС адаптації навчального контенту спеціальності «Кібербезпека» до вимог ринку праці підтвердила достатньо високу достовірність віднесення сформованих респондентами векторів ознак до відповідних класів розпізнавання. При цьому відкривається шлях для подальшого підвищення функціональної ефективності ІАС шляхом збільшення глибини машинного навчання й оптимізації додаткових параметрів функціонування системи.

Література

1. Агапова, М. О. Про проблему оптимізації навчального матеріалу [Текст] / М. О. Агапова // Теорія і практика управління соціальними системами. – 2012. – № 1. – С. 34–39.

2. Тецький А. Г. Аспекти кібербезпеки платформ дистанційного навчання / А. Г. Тецький, О. І. Морозова // Радіоелектронні і комп'ютерні системи. – 2020. – № 4(96). – С. 93-97. DOI: 10.32620/reks.2020.4.08.

3. Sklyar, V. V. *Case-Based and Project-Based Methods for Effective E-learning in ICT Safety and Security [Text]* / V. V. Sklyar, V. S. Kharchenko // *ICTERI Workshops 2020*. – P. 157-170.

4. Довбиш, А. С. Інтелектуальні інформаційні технології в електронному навчанні [Текст] /

А. С. Довбиш, А. В. Васильєв, В. О. Любчак. – Суми : Видавництво Сумського державного університету, 2013. – 178 с.

5. Xu, G. *Applied Data Mining [Text]* / G. Xu, Y. Zong, Y. Z. Yang. – CRC Press, 2013. – 284 p.

6. Hastie, T. *The Elements of Statistical Learning data Mining, Inference and Prediction [Text]* / T. Hastie, R. Tibshirami, J. Eridman. – 2nd ed. – Springer Verlag, 2009. – 746 p.

7. Patel, H. A. *Applications of Artificial Neural Networks for Nonlinear Data [Text]* / H. A. Patel, A. V. S. Kumar. – *Engineering Science Reference*, 2020. – 300 p.

8. Jayaswal, P. *Development of EBP-Artificial neural network expert system for rolling element bearing fault diagnosis [Text]* / P. Jayaswal, S. N. Verma, A. K. Wadhvani // *Journal of Vibration and Control*. – 2010. – Vol. 17, No. 8. – P. 1131-1148. DOI: 10.1177/1077546310361858.

9. Subbotin, S. *The neuro-fuzzy network synthesis and simplification on precedents in problems of diagnosis and pattern recognition [Text]* / S. Subbotin // *Optical Memory and Neural Networks (Information Optics)*. – 2013. – Vol. 22. – P. 97-103. DOI: 10.3103/s1060992x13020082.

10. Аспекти кібербезпеки платформ дистанційного навчання [Текст] / І. В. Шелехов, Н. Л. Барченко, В. В. Кальченко, В. К. Ободяк // Радіоелектронні і комп'ютерні системи. – 2020. – № 4(96). – С. 106-115. DOI: 10.32620/reks.2020.4.10.

11. *Information-Extreme Machine Learning of On-Board Vehicle Recognition System [Text]* / A. S. Dovbysh, M. M. Budnyk, V. Yu. Piatachenko, M. I. Myronenko // *Cybernetics and Systems Analysis*. – 2020. – No. 56. – P. 534-543. DOI: 10.1007/s10559-020-00269-y.

12. Naumenko, I. *Information-extreme learning of on-board system for recognition of ground vehicle [Text]* / I. Naumenko, M. Myronenko, V. Piatachenko // *Proceedings of the Second International Workshop on Computer Modeling and Intelligent Systems (CMIS-2019)*, Corpus ID: 146120351. – P. 121-132.

13. *Functional Diagnostic System for Multichannel Mine Lifting Machine Working in Factor Cluster Analysis Mode [Text]* / V. I. Zimovets, S. V. Shamatin, D. E. Olada, N. I. Kalashnykova // *Journal of Engineering Sciences*. – 2020. – Vol. 7, Iss. 1. – P. E20-E27. DOI: 10.21272/jes.2020.7(1).e4.

14. Schonlau, Matthias. *Visualizing non-hierarchical and hierarchical cluster analyses with clustergrams [Text]* / Matthias Schonlau // *Computational Statistics*. – 2004. – Vol. 19. – P. 95-111. DOI: 10.1007/BF02915278.

References

1. Agapova, M. O. Pro optymizaciju navchal'nogo material [About optimization of educational material]. *Teorija i praktyka upravlinnja social'nyu systematy* –

Theory and practice of social systems management, 2012, no. 1, pp. 34-39.

2. Tetskyi, A., Morozova, O. Aspekty kiberbezpeky platform dystantsiinoho navchannia [Cybersecurity aspects of e-learning platforms]. *Radioelektronni i komp'uterni sistemi – Radioelectronic and computer systems*, 2020, vol. 4(96), pp. 93-97. DOI: 10.32620/reks.2020.4.08.

3. Sklyar, V. V., Kharchenko, V. S. Case-Based and Project-Based Methods for Effective E-learning in ICT Safety and Security. *ICTERI Workshops 2020*, pp. 157-170.

4. Dovbysh, A. S., Vasylev, A. V., Lubchak, V. O. *Intelektual'ni informacijni tehnologii' v elektronnomu navchanni* [Intellectual information technologies in electronic learning]. Sumy, Sumy State University Publ., 2013. 172 p.

5. Xu, G., Zong, Y., Yang, Y. Z. *Applied Data Mining*. CRC Press Publ., 2013. 284 p.

6. Hastie, T., Tibshirami, R., Eridman, J. *The Elements of Statistical Learning: data Mining*. Inference and Prediction, Springer Verlag Publ., 2009. 746 p.

7. Patel, H. A., Kumar, A. V. S. *Applications of Artificial Neural Networks for Nonlinear Data*. Engineering Science Reference, 2020. 300 p.

8. Jayaswal, P., Verma, S. N., Wadhvani, A. K. Development of EBP-Artificial neural network expert system for rolling element bearing fault diagnosis. *Journal of Vibration and Control*, 2010, vol. 17, no. 8, pp. 1131-1148. DOI: 10.1177/1077546310361858.

9. Subbotin, S. The neuro-fuzzy network synthesis and simplification on precedents in problems of diagnosis and pattern recognition. *Optical Memory and*

Neural Networks (Information Optics), 2013, vol. 22, pp. 97-103. DOI: 10.3103/s1060992x13020082.

10. Shelechov, I., Barchenko, N., Kalchenko, V., Obodiak, V. Nechitka iierarkhichna otsinka yakostikompleksnykh system zakhystu informatsii [A hierarchical fuzzy quality assessment of complex security information systems]. *Radioelektronni i komp'uterni sistemi – Radioelectronic and computer systems*, 2020, vol. 4(96), pp. 106-115. DOI: 10.32620/reks.2020.4.10.

11. Dovbysh, A. S., Budnyk, M. M., Piatachenko, V. Yu., Myronenko, M. I. Information-Extreme Machine Learning of On-Board Vehicle Recognition System. *Cybernetics and Systems Analysis*, no. 56, pp. 534-543. DOI: 10.1007/s10559-020-00269-y.

12. Naumenko, I., Myronenko, M., Piatachenko, V. Information-extreme learning of on-board system for recognition of ground vehicle. *Proceedings of the Second International Workshop on Computer Modeling and Intelligent Systems (CMIS-2019)*, Corpus ID: 146120351, pp. 121-132.

13. Zimovets, V. I., Shamatin, S. V., Olada, D. E., Kalashnykova, N. I. Functional Diagnostic System for Multichannel Mine Lifting Machine Working in Factor Cluster Analysis Mode. *Journal of Engineering Sciences*, 2020, vol. 7, iss. 1, pp. E20-E27. DOI: 10.21272/jes.2020.7(1).e4.

14. Schonlau, Matthias. Visualizing non-hierarchical and hierarchical cluster analyses with clustergrams. *Computational Statistics*, 2004, vol. 19, pp. 95-111. DOI: 10.1007/BF02915278.

Надійшла до редакції 11.01.2021, розглянута на редколегії 16.02.2021

ИНФОРМАЦИОННО-АНАЛИТИЧЕСКАЯ СИСТЕМА ОЦЕНКИ СООТВЕТСТВИЯ УЧЕБНОГО КОНТЕНТА СПЕЦИАЛЬНОСТИ КИБЕРБЕЗОПАСНОСТЬ СОВРЕМЕННЫМ ТРЕБОВАНИЯМ

А. С. Довбыш, И. В. Шелехов, Ю. А. Хибовская, А. В. Матяш

Решена актуальная задача повышения функциональной эффективности машинного обучения информационно-аналитической системы (ИАС) оценки соответствия современным требованиям контента учебных дисциплин бакалаврского уровня специальности «Кибербезопасность». Разработано метод информационно-экстремального машинного обучения ИАС с целью адаптации учебного контента выпускающей кафедры к требованиям рынка труда, что позволяет при функционировании системы в режиме мониторинга оперативно корректировать контент учебных дисциплин выпускающей кафедры. Идея метода состоит в максимизации информационной способности ИАС в процессе машинного обучения, что позволяет в режиме мониторинга достигнуть максимальную полную вероятность принятия правильных классификационных решений. В результате информационно-экстремального машинного обучения в рамках геометрического подхода построены решающие правила, практически инвариантные к многомерности пространства признаков распознавания. В качестве критерия оптимизации параметров машинного обучения используется модификация информационной меры Кульбака, являющаяся функционалом точностных характеристик классификационных решений. В качестве параметров оптимизации рассматривались геометрические параметры гиперсферических контейнеров классов распознавания, которые в процессе машинного обучения восстанавливались в радиальном базисе бинарного пространства признаков Хемминга. При этом входная обучающая матрица трансформировалась в рабочую бинарную обучающую матрицу, изменяющуюся в процессе машинного обучения путём допустимых преобразований с целью адаптации входного математического описания системы к максимальной достоверности классификационных решений. Предложено категорийную модель функционирования ИАС, на основе которой разработан алгоритм информационно-экстремального машинного обучения системы с автоматическим определением базового

класа розпознавання. По результатам опроса специалистов в области кибербезопасности сформировано входную структурированную обучающую матрицу і по результатам физического моделирования подтверждено работоспособность предложенного метода информационно-экстремального машинного обучения ИАС.

Ключевые слова: информационно-экстремальная интеллектуальная технология; машинное обучение; информационный критерий; оптимизация; учебный контент; базовый класс распознавания; кибербезопасность.

INFORMATION AND ANALYTICAL SYSTEM FOR ASSESSING THE COMPLIANCE OF EDUCATIONAL CONTENT SPECIALTIES CIBER SECURITY WITH MODERN REQUIREMENTS

A. Dovbysh, I. Shelekhov, Ju. Khibovska, O. Matiash

The urgent task of increasing the functional efficiency of machine learning of the information and analytical system (IAS) for assessing compliance with modern requirements of the bachelor level academic disciplines content of the specialty "Cybersecurity" has been solved. A method of information-extremal machine learning IAS has been developed to adapt the educational content of the graduating department to the requirements of the labor market, which allows, when the system is operating in the monitoring mode, to promptly correct the content of the graduating department educational disciplines. The idea of the method is to maximize the information capacity of the IAS in the process of machine learning, which makes it possible to achieve the maximum total probability of making correct classification decisions in the monitoring mode. Because of information-extremal machine learning in the framework of the geometric approach, decisive rules have been constructed that are invariant to the multidimensionality of the recognition feature space. As a criterion for optimizing machine learning parameters, a modification of the Kullback information measure is used, which is a function of the accuracy characteristics of classification decisions. As optimization parameters, the geometric parameters of hyperspherical containers of recognition classes were considered, which were restored in the process of machine learning in the radial basis of the binary space of Hamming features. Simultaneously, the input-training matrix was transformed into a working binary-training matrix, which changes in the process of machine learning by using admissible transformations to adapt the input mathematical description of the system to the maximum reliability of classification solutions. A categorical model of IAS functioning is proposed, based on which an algorithm for information-extreme machine learning of the system with automatic determination of the recognition base class is developed. According to the results of a specialist survey in the field of cybersecurity, an input-structured training matrix was formed and based on the results of physical modeling, the efficiency of the proposed method of information-extreme machine learning IAS was confirmed.

Keywords: information-extreme intelligent technology; machine learning; information criterion; optimization; educational content; basic recognition class; cybersecurity.

Довбиш Анатолій Степанович – д-р техн. наук, професор, завідувач кафедри комп'ютерних наук Сумського державного університету, Суми, Україна.

Шелехов Ігор Володимирович – канд. техн. наук, доцент, доцент кафедри комп'ютерних наук Сумського державного університету, Суми; докторант кафедри комп'ютерних систем, мереж і кібербезпеки, Національний аерокосмічний університет ім. М. Є. Жуковського «Харківський авіаційний інститут», Харків, Україна.

Хібовська Юлія Олексіївна – провідний фахівець кафедри комп'ютерних наук Сумського державного університету, Суми, Україна.

Матяш Олександр Вячеславович – аспірант Сумського державного університету, Суми, Україна.

Anatoliy Dovbysh – Doctor of Technical Science, Head of the Computer Science Department of Sumy State University, Sumy, Ukraine,

e-mail: a.dovbysh@cs.sumdu.edu.ua, ORCID: 0000-0003-1829-3318, Scopus Author ID: 36052468600,

ResearcherID: AAH-1630-2021, <https://scholar.google.com/citations?user=AcBgyaAAAAAJ&hl=ru>

Igor Shelekhov – PhD, Associate Professor at the Computer Science Department, Sumy State University, Sumy; DrS-student at the Department of Computer Systems, Networks and Cybersecurity, National Aerospace University "Kharkiv Aviation Institute", Kharkiv, Ukraine,

e-mail: i.shelekhov@cs.sumdu.edu.ua, i.shelekhov@csn.khai.edu, ORCID: 0000-0003-4304-7768,

Scopus Author ID: 55537177800, ResearcherID: AAD-4757-2019,

https://scholar.google.com/citations?user=84_CNroAAAAAJ

Yuliya Khibovska – Senior Specialist of the Computer Science Department of Sumy State University, Sumy, Ukraine,

e-mail: y.khibovska@cs.sumdu.edu.ua, ORCID: 0000-0001-5832-3134.

Oleksandr Matiash – PhD Student of Sumy State University, Sumy, Ukraine,

e-mail: alex.matiash97@gmail.com, ORCID: 0000-0002-2020-1614.