

DOI [10.28925/2663-4023.2022.18.108123](https://doi.org/10.28925/2663-4023.2022.18.108123)

УДК 336.71:004.056

Лукова-Чуйко Наталія Вікторівна

доктор технічних наук, професор,
завідувачка кафедри кібербезпеки та захисту інформації
Факультету інформаційних технологій
Київський національний університет імені Тараса Шевченка, Київ, Україна
ORCID ID: 0000-0003-3224-4061
lukova@knu.ua

Лаптева Тетяна Олександрівна

аспірантка кафедри кібербезпеки та захисту інформації
Факультет інформаційних технологій
Київський національний університет імені Тараса Шевченка, Київ, Україна
ORCID ID: 0000-0002-5223-9078
tetiana1986@ukr.net

МЕТОД РОЗРОБКИ КЛАСИФІКАТОРА ЗІ ЗАСТОСУВАННЯМ ТЕОРЕМИ БАЙЄСА (BAYES) ДЛЯ УХВАЛЕННЯ РІШЕННЯ ПРО ВИЗНАЧЕННЯ ПРАВДИВОЇ ІНФОРМАЦІЇ

Анотація. Спектр застосування кластерного аналізу дуже широкий: його використовують в археології, медицині, психології, біології, державному управлінні, регіональній економіці, маркетингу, соціології та інших дисциплінах. Кожна дисципліна висуває свої вимоги до первинних даних і правил формування груп. Очевидно, різними будуть методологічні підходи до сегментації ринку, мета якого визначити групи об'єктів, схожих між собою за ознаками та властивостями і до формування кластерів, які об'єднуються задля посилення своїх конкурентних переваг. Так при обробки інформації в інформаційному просторі методологія спрямована, зазвичай, на побудову математичної моделі кластерного аналізу досліджуваного об'єкта чи явища, і навіть отримання відповіді на питання: «Чи правдива інформація чи ні». Виявлення недостовірної інформації у цифровому світі є важливим завданням у подоланні широкого розповсюдження чуток та упереджень.

В роботі проведено аналіз існуючих методів класифікації інформації у інформаційної епохи. Сформулювати ознаки інформаційної епохи, в контексті визначення правдивості інформації. На базі основних ознак інформаційної епохи розроблено метод створення класифікатора для вирішення завдань визначення правдивості інформації.

Проведено математичне моделювання з використанням розробленого класифікатора для підтвердження розробленого методу ухвалення рішення о правдивості інформації за допомогою теореми Байєса (Bayes). Отримані результати довели працездатність запропонованого методу розробки класифікатора для якого при застосуванні теореми Байєса (Bayes) для ухвалення рішення можливо визначити правдивість інформації.

Але розроблений байєсовський класифікатор ґрунтується на тому, що апіорні ймовірності гіпотез - відомі. Тому напрямком подальших досліджень є розробка або удосконалення методів та алгоритмів визначення апіорні ймовірності гіпотез.

Ключові слова: класифікатор; кластеризація; простір образів; теорема Байєса; правдивість інформації, апіорна ймовірність.

ВСТУП

Наслідками стрімкого розвитку інформаційних технологій у світі та Україні є суттєве прискорення інформатизації суспільної діяльності, посилення процесів глобалізації та інтернаціоналізації. Активізація інформаційних процесів вплинула й на систему стратегічних комунікацій суспільства, у якій провідна роль відводиться

комунікативній складовій. Комунікативна складова є невідкладною частиною інформаційної системи.

Під інформаційною системою розуміється система, яка здійснює:

- одержання вхідних даних;
- обробку цих даних та/або зміна власної внутрішнього стану (внутрішніх зв'язків/відносин);
- видачу результату або зміна свого зовнішнього стану;(зовнішніх зв'язків/відносин).

Інформаційну систему, елементи якої функціонують у відповідно до правил, породжених одним і тим же взаємно несуперечливим безліччю аксіом, назвемо простою інформаційною системою.

Інформаційну систему, яка містить елементи, що функціонують відповідно до правил, породжених відмінними один від одного множинами аксіом, назвемо складною інформаційною системою.

Завдяки стрімкому розвитку інформаційних технологій, збереження та використання інформації з метою задоволення інформаційних інтересів і потреб громадян, інформаційні системи представляють собою невід'ємну складову національного інформаційного простору. Цей величезний обсяг інформації постав під загрозу правдивість новин, які поширюються. Підроблені новини – це будь-яка форма помилкової інформації чи контенту, що поширюється в інформаційній мережі, для впливу на погляд людей на певну подію чи інформацію.

Виявлення недостовірної інформації у цифровому світі є важливим завданням у подоланні широкого розповсюдження чуток та упереджень. Багато досліджень було проведено для виявлення елементів дезінформації для англійської мови, проте українська мова не має досліджень у цій галузі.

Постановка проблеми. Основне протиріччя, яке лежить в основі наукового дослідження полягає, в тому, що незважаючи на велику кількість публікацій, завдання виявлення неправдивої інформації у великій кількості інформаційного простору, залишається невирішеним. Це обумовлено постійною модифікацій методів та методик розповсюдження неправдивої інформації. Тому розробка та удосконалення методик прийняття рішення о правдивості інформації є актуальним науковим завданням. Цьому науковому завданню і присвячена дана робота.

Аналіз останніх досліджень і публікацій. Проблемі розробки класифікаторів для аналізу даних та ухвалення рішення на основі розробленого класифікатора про віднесення даних до того чи іншого процесу присвячено багато наукових робіт. На даний час в літературі описано багато прикладних задач, що вирішуються шляхом застосування методів кластерного аналізу.

Так у роботі [1] розглядається питання кластеризації, або кластерного аналізу, завдання якої полягає в розбитті вибірки об'єктів на підмножини, що не перетинаються та називаються кластерами. Кожен кластер має складається зі схожих об'єктів, а об'єкти різних кластерів мають істотно відрізнятися один від одного. Задача кластеризації відноситься до статистичної обробки, а також до широкого класу задач навчання без вчителя. Ще її можна описати через задачу класифікації. Але сам опис завдання класифікації не наводиться, питання побудови самого класифікатора не розглядаються.

У роботі [2] автори розглядають класифікацію як задачу аналізу об'єктів. Тобто до якого з відомих класів відносять досліджувані об'єкти, як їх класифікувати. Задачу класифікації розглядають як задачу визначення значення одного з параметрів



аналізованого об'єкта на підставі значень інших параметрів. Завдання створення методу побудови класифікатора не розглядають.

У роботах [3 – 4] визначають сутність кластерів, але обґрунтовано власне бачення того, за якими ознаками та яким чином повинна проводитися їх класифікація. Так, наприклад, у доцільно обґрунтовано доцільність створення та проведення класифікації взаємодіючих підприємств на основі пізнання закономірностей інтеграційних процесів. Наводяться ознаки об'єктів для класифікації, але процес визначення критерія відбору об'єктів для прийняття рішення для подальшої класифікації не наводиться.

У роботі [5] розглядається процес при котрому на підставі навчальної вибірки будується модель визначення значення залежної змінної, яку часто називають функцією класифікації або регресії. Для отримання максимально точної функції до навчальної вибірки пред'являються вимоги (про кількість об'єктів, що входять до вибірки, має бути досить великою; до вибірки мають входити об'єкти, що представляють усі можливі класи у задачі класифікації або всю область значень у задачі регресії; для кожного класу в задачі класифікації або кожного інтервалу області значень у задачі регресії вибірка має містити достатню кількість об'єктів.) але другий етап застосування якісь критеріїв до об'єктів аналізу не розглядається. Критеріїв оцінки об'єктів класифікатора не наводиться.

У роботах [6 – 7] говориться про те, що кластерний аналіз (або автоматична класифікація, розпізнавання образів без вчителя) займає одне з центральних місць серед методів аналізу даних і є сукупністю підходів, методів і алгоритмів, призначених для знаходження деякого розбиття досліджуваної сукупності об'єктів на підмножини схожих між собою об'єктів. При цьому вихідним припущенням для виділення таких підмножин, що отримали спеціальну назву кластерів, служить лише неформальне припущення про те, що об'єкти, які відносяться до одного кластера, повинні мати більшу схожість між собою, чим з об'єктами з інших кластерів. Конкретних методів та алгоритмів побудови класифікаторів та прийняття рішень на базі отриманих значень параметрів об'єктів класифікатора не наводиться.

Разом з тим, незважаючи на значну кількість публікацій щодо вирішення різноманітних аспектів оцінювання правдивості інформації на сьогоднішній день залишається невирішеною задача оцінювання правдивості інформації з урахуванням можливого інформаційного впливу на суспільство, ресурси. Тому проблема розробки методу створення класифікатора отриманої інформації та ухвалення рішення на основі розробленого класифікатора про визначення правдивої або неправдивої інформації є актуальним науковим завданням.

Мета статті. Провести аналіз існуючих методів класифікації інформації. Сформулювати ознаки інформаційної епохи, в контексті визначення правдивості інформації. На базі основних ознак інформаційної епохи розробити метод створення класифікатора для вирішення завдань визначення правдивості інформації. На базі розробленого класифікатора за допомогою теореми Байєса (Bayes) розробити метод ухвалення рішення о правдивості інформації.

РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Інформація завжди була, є і буде найважливішим із комунікативних ресурсів. Інформація є необхідною складовою функціонування усіх соціальних систем. У приватному житті, для управління складними технологічними системами або для



розбудови незалежної держави завжди є нагальна потреба у надійній та оперативній інформації. З метою задоволення інформаційних потреб громадян, юридичних осіб і держави органи державної влади та органи місцевого самоврядування здійснюють інформаційну діяльність та створюють інформаційні служби, системи, мережі, бази і банки даних.

Ознаки інформаційної епохи

1. Схема передачі інформації та знання за останні десятиліття зазнала серйозних змін, від схеми: «людина – людина» до схеми: «людина – технічний засіб – людина».

2. Процеси виробництва, поширення, впровадження інформації та оцінки результатів інформаційного впливу стали носити промисловий характер, вони поставлені на конвеєр. Відбувся перехід від схеми: «людина – інформаційний продукт» до схеми: «людина – технічний засіб – інформаційний продукт».

3. Сучасна інформаційна зброя за критерієм ефективність/вартість значно перевершує будь-який інший вид озброєння.

4. Делегування владних повноважень у країнах, що визначають новий світовий порядок, здійснюється на базі інформаційних технологій.

5. У межах ареалу поширення людства передача інформації з допомогою технічних засобів (перепрограмування собі подібних, використовуючи сучасні інформаційні технології) здійснюється значно швидше, ніж передача генетичної інформації (програмування своїх нащадків, використовуючи біологічні механізми). Понад те, обсяги й час передачі у межах ареалу поширення людства основної маси еліти, визначальною функціонування системи управління держав і народів, перестали залежати від місця розташування у просторі, від відстані з-поміж них.

Процес передачі інформації від офіційного джерела чи з “місця події” до кінцевого споживача є достатньо складним і тривалим. Під час свого просування інформація циркулює в інформаційному просторі і перебуває під впливом різних груп впливу, які переслідують власні інтереси. Отже, дуже часто кінцевий користувач отримує упереджену, необ’єктивну інформацію, метою якої є вчинення певного впливу на його поведінку. Однією з проблем, яка стримує впровадження ефективних систем захисту інформаційних ресурсів організації чи держави, є проблема створення достовірної класифікації атак та механізмів «фільтрування» інформації, яка проходить від першоджерела до споживача. Виявлення та відсіч неправдивої інформації.

Розробка методу створення класифікатора для вирішення завдань визначення правдивості інформації.

Для розробки моделі потрібно встановити обмеження та застосувати початкові дані, далі на основі моделі буде розроблено метод методу створення класифікатора для котрого застосуємо теорему Байєса (Bayes) для ухвалення рішення про визначення правдивої інформації. Головна мета це ухвалення рішення про визначення правдивої інформації.

Початкові дані для моделі розробки методу визначення правдивості інформації :

- два та більше суб'єктів, які здійснюють боротьбу один з одним за ресурси;
- поведінка суб'єктів, що борються, визначається наявними у них моделями світу. При цьому моделі світу зазнають постійних змін під впливом вхідних даних, що надходять на вхід систем;
- боротьба здійснюється лише шляхом цілеспрямованого інформаційного впливу, тобто, завдяки створенню відповідних послідовностей повідомлень, вкладених у спотворення моделі світу противників.

Результатом вирішення названих завдань мають стати моделі та стратегії поведінки суб'єктів (інформаційних об'єктів).

Для розробки моделі будемо використовувати лінійний курс алгебри, теорію ймовірностей, методи, які називають «м'якими обчисленнями». Методи – «м'якого обчислення» асоціюються з інтелектуальними завданнями, це пов'язано, з тим, що вони мають високу універсальність і можуть бути застосовані до того, як буде розроблено хороший спеціалізований метод вирішення завдання. Методи – «м'якого обчислення» – сукупність емпіричних, нечітких, наближених методів вирішення завдань, які не мають точних алгоритмів рішення за поліноміальний час (час роботи яких поліноміальне залежить від обсягу вхідних даних). Іншою важливою загальною властивістю м'яких обчислень є властивість адаптивності – підстроювання під завдання. Це вносить у розв'язання задачі етап навчання. Можливість працювати із завданням як із чорним (система з невідомим алгоритмом, про суть якого можна тільки здогадуватися за зовнішніми проявами) або сірою скринькою (система з можливістю обмеженого налаштування користувачем), не вникаючи до кінця у тонкощі роботи керованої системи, сподіваючись, що ці тонкощі будуть компенсовані «інтелектом» методу. У сучасній науці, до м'яких обчислень прийнято відносити такі методи: нейронні мережі, еволюційні стратегії, нечітка логіка, багато агентні системи (інтелект зграї), різні евристичні алгоритми пошуку рішень та ін.

Зі збільшенням обсягів інформації, що обробляється, зберігається та накопюється в результаті роботи інформаційних систем і процесів, в ході діяльності підприємств або науково-дослідної діяльності, її обробка і аналіз стають скрутним. Таким чином, виникає необхідність первісної обробки інформації для її структурування, виділення характерних ознак, узагальнення, сортування. На рис.1. наведено схема обробки інформації



Рис.1. Етапи обробки інформації



1. Пошук інформації. На першому кроці нам потрібно знайти, які документи повинні бути проаналізовані і забезпечити доступ до них. Зазвичай, користувачі можуть вибрати документи, що проаналізуються самі – тобто вручну, але якщо кількість документів буде велика, то необхідно використовувати варіант автоматизованого збору за певними критеріями.

2. Попередня обробка документів. На цьому кроці виконуються необхідні перетворення з документами для того щоб показати їх у вигляді, з яким можуть працювати математичні методи. Потрібні ці перетворення для видалення зайвої інформації.

3. Витягнення інформації. Витяг інформації з наших документів означає відображення в них ключових понять, з якими далі будемо виконувати аналіз.

4. Застосування методів аналізу. На цьому етапі витягуються шаблони і відношення, які є в інформації. Це основний крок у процесі аналізу інформації, і практичні задачі, що вирішуються.

5. Інтерпретація результатів. Останній крок у процесі отримання інформації передбачає інтерпретацію отриманих результатів. Зазвичай, інтерпретація полягає в їх візуалізації в графічному виді.

Саме для процесу аналізу застосовують процеси класифікації та кластеризації, що дозволяють проводити первинну обробку інформації, на її подальшого аналізу. Практично всі методи класифікації та кластеризації засновані на гіпотезі компактності. Її можна сформулювати таким чином – об'єкти, що належать до одного класу, повинні бути розташовані компактно хоча б в одному з можливих просторів опису об'єкта.

Кластеризація — процес розбиття заданої вибірки об'єктів (спокус) на підмножини (зазвичай, непересічні), звані кластерами, те щоб кожен кластер складався з подібних об'єктів, а об'єкти різних кластерів істотно відрізнялися.

Однією з цілей кластеризації є виявлення внутрішніх зв'язків між даними шляхом визначення кластерної структури. Розбиття спостережень на групи подібних об'єктів дозволяє спростити подальшу обробку даних та прийняття рішень, застосовуючи до кожного кластера свій метод аналізу - "divide et impera" (стратегія "розділай і володарюй"). Одним із додатків кластеризації є вирішення задачі стиснення даних. Якщо вихідна вибірка надмірно велика, можна скоротити її, залишивши кілька найбільш характерних представників від кожного кластера.

Класифікація — процес упорядкування або розподілу об'єктів. (спостережень) за класами з метою відображення відносин між ними.

Клас — це безліч об'єктів, що мають певну загальну ознаку, що відрізняє цю сукупність з інших об'єктів. Як класифікаційного поділу можна прийняти різні ознаки, що залежать від мети класифікації. В основу класу завжди кладуть найважливішу ознаку об'єкта, що відповідає меті класифікації.

Класифікувати об'єкт означає вказати клас, до якого належить цей об'єкт. Результатом класифікації об'єкта є найменування класу, що визначається алгоритмом класифікації в результаті його застосування до даного конкретного об'єкта.

Навчання класифікатора - процес настройки алгоритму у разі, коли задано кінцеву множину об'єктів, для яких відомо, до яких класів вони відносяться. Ця множина називається вибіркою. Приналежність до того чи іншого класу інших об'єктів невідома. Реалізація завдання класифікації спочатку повинна спиратися на задані постулати, основні з яких – апріорна інформація про первинну множину об'єктів та міру близькості елементів та класів.

**Постановка задачі класифікації для виявлення неправдивої інформації.**

Для визначення класифікатора для виявлення неправдивої інформації якій будемо розробляти, введемо визначення та обмеження:

Нехай Ω – безліч об'єктів розпізнавання (простір образів).

$\omega \in \Omega$ об'єкт розпізнавання (образ).

$g(\omega): \Omega \rightarrow I, I = \{1, 2, \dots, n\}$ - індикаторна функція, що розбиває простір образів Ω на n класів, що не перетинаються $\Omega^1, \Omega^2, \dots, \Omega^n$. Індикаторна функція невідома спостерігачеві.

X – простір спостережень, що сприймаються спостерігачем (простір ознак).

$x(\omega): \Omega \rightarrow X$ — функція, що ставить у відповідність кожному об'єкту ω точку $x(\omega)$ у просторі ознак.

$x(\omega)$ – це образ об'єкта, який сприймає спостерігач.

У просторі ознак визначені непересічні множини точок $[i] \subset X, i=1, 2, \dots, n$, відповідних образам одного класу.

$\varphi(x): X \rightarrow I$ вирішальне правило – оцінка для $g(x)$ виходячи з $x(\omega)$, тобто.

$\varphi(x) = \varphi(x(\omega))$.

Нехай $x_v = x(\omega_v), v=1, 2, \dots, N$ доступна спостерігачеві інформація про функції $g(\omega)$ та $x(\omega)$, але самі ці функції спостерігачеві невідомі. Тоді $(g_v, x_v), v=1, 2, \dots, N$ — є безліч прецедентів.

Завдання полягає у побудові такого вирішального правила $\varphi(x)$, щоб розпізнавання проводилося з мінімальним числом помилок.

Звичайний випадок - *вважати простір ознак евклідовим, а якість вирішального правила вимірювати частотою появи правильних рішень.*

Як правило, його оцінюють, наділяючи безліч об'єктів деякою імовірнісною мірою. Байєсовський підхід виходить із статистичної природи спостережень. За основу береться *припущення про існування ймовірнісної міри на просторі образів, яка відома, чи то, можливо оцінена. Мета полягає у розробці такого класифікатора, який буде правильно визначати найімовірніший клас для пробного образу. Тоді завдання полягає у визначенні “найвірогіднішого” класу.*

Байєсовський підхід ґрунтується на припущенні про існування деякого розподілу ймовірностей для кожного параметра.

Використання пошуку відповідності передусь побудова безлічі статистик, в яких міститься кількість текстів у даному класі та список термів, що використовуються разом зі своїми лічильниками.

Для визначення відповідного класу текстів для заданого тексту будується структура з неповторних термів та їх лічильників – $(w_i, n(w_i))$.

Через M позначимо розмір множини статистик. Класи, належність до яких перевіряється текст, позначимо через $c_j (j=0, 1, \dots, M-1)$. Для кожного слова w_i з тексту, що перевіряється, у кожній статистиці знаходимо це слово і відповідний лічильник $n(w_i, c_j)$, (тут $j (j=0, 1, \dots, M-1)$ -номер класу (елемент множини статистик)). Через $p(c_j)$ позначимо кількість текстів у j -му класі. Мінімізація ризику та ймовірності помилки еквівалентні поділу простору ознак на n областей. Якщо області суміжні, всі вони розділені поверхнею рішення багатомірному просторі. Для випадку побудови поверхні, що розділяє, краще використовувати методи класифікації відмінні від Байєсовської. Використання імовірнісних характеристик визначається розподілом Гауса, який дуже широко використовується через обчислювальну зручність і адекватність у багатьох випадках.

Якщо відомо, або з достатньою підставою можна вважати, що щільність розподілу функцій правдоподібності $P(x|\Omega_i)$ є гауссівською, то застосування класифікатора Байєса призводить до того, що образи, що характеризуються нормальним розподілом виявляють тенденцію до групування навколо середнього значення, які розсіювання пропорційне середньоквадратичному відхиленню σ . Імовірнісні методи спираються на інформацію про густину розподілу ймовірностей кожного класу. На жаль, у реальних завданнях інформація про густину розподілу відсутня.

Використання теореми Байєса (Bayes) для ухвалення рішення

Розглянемо один із найпопулярніших методів класифікації, заснований на стохастичних принципах. Основою цього методу є апіорна інформація про розподіл ймовірностей існуючих класів C_i .

Апіорна (вихідна) ймовірність інтерпретується як опис інформації без свідчення події. Апостеріорна (наступна) ймовірність приймає це свідчення до уваги. Апіорна ймовірність появи події у при спостереженні x обчислюється так:

$$P(y|x) = \frac{P(x|y)P(y)}{P(x)} \quad (1)$$

де $P(x)$ апіорна ймовірність спостереження x ;
 $P(y)$ апіорна ймовірність появи події y .

Розглянемо наступне завдання – щодо виявлення достовірності інформації, ухвалити рішення про те, що це достовірне чи ні.

Оцінка достовірності інформації – досить складний процес, якому треба вчитися. Це вміння відібрати безліч професійно значущих джерел інформації, необхідних для успішного розвитку певної сфери діяльності в кожен конкретний період часу.

За головне виділимо кілька кроків оцінки достовірності:

- Прочитайте текст: чи немає там зайвих повторюваних слів?
- Чи немає в тексті зайвої емоційності і сенсаційності?
- Перевірте текст на унікальність.
- Правило трьох джерел: інформація повинна повторюватися як мінімум у трьох незалежних джерелах.

Для аналізу інформації важливо вміти *розділяти факти і думки*. Це можна зробити за наступним алгоритмом:

- роздрукуйте текст;
- виділіть інформаційні факти і коментарі цих фактів;
- шукаємо інформацію по темі і проводимо аналіз, звертаючи увагу на наступні моменти:

- автор сам не володіє всією інформацією;
- автор навмисне приховує від споживача частину інформації;
- автор домислює інформацію;
- неправильне цитування;
- ермінологічна плутанина.

"Довіряй але перевірйй". Для оцінки достовірності інформації можна скористатися рекомендаціями:

- Перевірка фактичного матеріалу.
- Встановлено автентичність факту?
- Чи точно наведено факти і числа?



- Чи посилання на авторитетні джерела?

Порівняння - один з найефективніших способів пошуку істини. Рідко одна і та ж недостовірна інформація публікується на декількох сайтах відразу.

Встановлення використання матеріалу *іншими джерелами*. Передрук і копіювання даних з одного сайту іншими сайтами є хорошим знаком, оскільки це означає, що цьому джерелу довіряють.

З'ясування рейтингу і авторитету джерела. Відомі ресурси зазвичай заслуговують довіри, оскільки трепетно ставляться до свого рейтингу і не стануть розмінювати його на сумнівні сенсації.

Отримання інформації про автора матеріалу:

- Наскільки компетентний автор матеріалу?
- Що ще він написав?
- Чи відгуки читачів?

Авторитетність джерела. Це:

- авторитетність відносна, конкретна і контекстуально залежна;
- найбільш авторитетне джерело - оглядова стаття у поважному науковому журналі;
- книги, якщо вони видані в авторитетних видавництвах авторитетними авторами;

Бійтеся помилкової авторитетності. Не довіряйте сліпо твердженнями про авторитетність якого-небудь джерела. Вчені ступені означають авторитет тільки в тій області, в якій вони були отримані. Особливо в Інтернеті хто завгодно може опублікувати що завгодно

Критерії якості інформації

Якість інформації можна оцінити за наступними критеріями:

- Своєчасність інформації.
- Коли інформація опублікована?
- Інформація була переглянута або оновлена?
- Посилання функціональна?
- Значимість інформації.

Мета. Яка мета цієї інформації? Повідомити? Навчити? Продати? Розважити? Переконати?

Роблять автори свої наміри чи цілі ясними?

Дана інформація - факт, думка або пропаганда?

Викладена точка зору об'єктивна і неупереджена?

Чи політичні, ідеологічні, культурні, релігійні або особисті пристрасті?

Але використовувати усі ознаки не завжди доцільно тому з, щоб оцінити достовірність інформації в інформаційному просторі пропонуються наступні ознаки:

- сумнівність викладених фактів, що визначається приховуванням джерел і авторів контенту, недостатньою аргументацією, посиланнями на думку широкого загалу, наявністю риторичних запитань;
- емоційне забарвлення контенту, що використовується для відображення емоційного стану автора і проявляється у перенасиченні контенту образними засобами, прикметниками, порівняннями тощо;
- тональність контенту по відношенню до деякого об'єкту чи події, яка відображає оцінювальні судження автора і може проявлятися у використанні зображень тощо;
- сенсаційність контенту, яка має на меті привернути увагу завдяки підвищенню тривожності та ін..;
- прихований (імпліцитний) зміст контенту пов'язаний з його глибинним змістом.

На перерахованих вище ознаках і буде ґрунтуватися моніторинг достовірності інформації в інформаційному просторі. Для ухвалення рішення буде використовуватися теорема Байєса (Bayes).

Розглянемо завдання по ознакам які характерні інформації прийняти рішення про достовірність інформації.

Для цього нам потрібно розробити критерії нормування цих ознак. Пропонується використовувати нормування проводити виходячи з можливості отримання ознакою 100% гарантії відповідності правдивості інформації. Наприклад:

Мета

1. Яка мета цієї інформації? Повідомити? Навчити? Продати? Розважити? Переконати?
2. Роблять автори свої наміри чи цілі ясними?
3. Дана інформація - факт, думка або пропаганда?
4. Викладена точка зору об'єктивна і неупереджена?
5. Чи політичні, ідеологічні, культурні, релігійні або особисті пристрасті?

Тобто пропонується кожному питанню надати рівні оцінки по 20% (у нормованому вигляді 0,2), усі разом будуть давати 100% -1 відповідно у нормованому стані. Якщо коефіцієнт буде менш ніж 50% - у нормованому вигляді це буде менш чим 0,5 - тоді інформація буде вважатися неправдивою.

Виходячи з прийнятої гіпотези визначення оцінок для ознак, заповнимо таблицю для будь якої інформації – у довільних оцінках.

Таблиця 1.

Запропонований класифікатор для вирішення завдань визначення правдивості інформації.

№	Ознаки	Правдива	Неправдива
1	Сумнівність викладених фактів	0,55	0,3
2	Емоційне забарвлення контенту	0,7	0,2
3	Тональність контенту	0,65	0,35
4	Сенсаційність контенту	0,85	0,45
5	Прихований (імпліцитний) зміст контенту	0,55	0,45
6	З'ясування рейтингу і авторитету джерела	0,57	0,33
7	"Довіряй але перевірй"	0,55	0,5
8	Порівняння	0,75	0,25
9	Якість інформації	0,55	0,45
10	Мета	0,45	0,35
Середнє квадратичне відхилення σ^2		0,119	0,098
Середнє значення ознаки μ		0,55	0,4

Ми будемо враховувати що середній коефіцієнт для правдивої інформації буде 0,55. Для неправдивої 0,4.

Зробимо ще одне припущення, вважатимемо, що інформація у інформаційному просторі має нормальний розподіл $N(\mu, \sigma^2)$ закон Гаусса. Це припущення засновується на тому, що закон Гаусса є граничним законом, до якого наближаються (за певних умов) інші закони розподілу. Для нормованого нормального розподілу характерно, що в інтервал $\mu \pm \sigma$ потрапляють 68% усіх результатів, в інтервал $\mu \pm 2\sigma$ потрапляють 95% усіх результатів, в інтервал $\mu \pm 3\sigma$ потрапляють 99% всіх результатів. Вираз для нормального закону розподілу:

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \quad (2)$$

Тоді правдивість інформації має розподіл:

$$p(x) = \frac{1}{0,119 \cdot \sqrt{2\pi}} \exp\left[-\frac{(x-0,55)^2}{2 \cdot (0,119)^2}\right] \quad (3)$$

Тоді неправдивість інформації має розподіл:

$$p(x) = \frac{1}{0,098 \cdot \sqrt{2\pi}} \exp\left[-\frac{(x-0,40)^2}{2 \cdot (0,098)^2}\right] \quad (4)$$

Необхідно оцінити параметри цих розподілів для їх використання у теоремі Байеса для прийняття рішення о правдивості чи неправдивості інформації. Для цього необхідно побудувати класифікатор максимальної правдоподібності. Він засновано на тому, що пріоритет відається той події, у якій найбільша ймовірність у даному випадку, тобто виконується нерівність: $p(x|n) > p(x|\bar{n})$ тобто інформація правдива якщо виконується нерівність:

$$\frac{1}{0,119 \cdot \sqrt{2\pi}} \exp\left[-\frac{(x-0,55)^2}{2 \cdot (0,119)^2}\right] > \frac{1}{0,098 \cdot \sqrt{2\pi}} \exp\left[-\frac{(x-0,40)^2}{2 \cdot (0,098)^2}\right] \quad (5)$$

У протилежному випадку інформація є неправдивою.

Апріорна ймовірність для правдивої та неправдивої інформації різна. Тому приймемо за припущення, що ймовірність того, що інформація неправдива $P_{\bar{n}}=3/5$, а то що правдива $P_n=2/5$, тоді у відповідності до правила Байеса:

для правдивої інформації $P(n|x) = [p(x|n) P_n] / p(x)$,

для неправдивої інформації та $P(\bar{n}|x) = [p(x|\bar{n}) P_{\bar{n}}] / p(x)$,

У цьому випадку класифікатор буде мати вигляд тільки якщо виконується нерівність:

$$P(n|x) = [p(x|n) P_n] / p(x) > P(\bar{n}|x) = [p(x|\bar{n}) P_{\bar{n}}] / p(x),$$

Або у іншому вигляді:

$$P(n|x) = [p(x|n) P_n] > P(\bar{n}|x) = [p(x|\bar{n}) P_{\bar{n}}],$$

Таким чином нерівність прийме вигляд:

$$\frac{3}{5} \cdot \frac{1}{0,119 \cdot \sqrt{2\pi}} \exp\left[-\frac{(x-0,55)^2}{2 \cdot (0,119)^2}\right] > \frac{2}{5} \cdot \frac{1}{0,098 \cdot \sqrt{2\pi}} \exp\left[-\frac{(x-0,40)^2}{2 \cdot (0,098)^2}\right] \quad (6)$$

Після рішення нерівності, отримуємо, якщо $x > 0,88$, то дана інформація правдива. Якщо $x < 0,88$ то інформація неправдива.

Для перевірки працездатності класифікатора, проведемо розрахунки згідно розробленої моделі. Для цього будемо змінювати апіорну ймовірність правдивості від 70 % до 30 % , а апіорну ймовірність неправдивої інформації від 30% до 70%. Відповідно апіорна ймовірність для правдивої інформації буде $P_p = \{0,8; 0,7; 0,6; 0,4; 0,3; 0,2\}$, а для неправдивої $P_n = \{0,2; 0,3; 0,4; 0,6; 0,7; 0,8\}$ відповідно.

Згідно запропонованого методу розробки класифікатора з використанням теореми Байєса (Bayes) для ухвалення рішення про визначення правдивої інформації розрахуємо параметри за якими можливо прийняти рішення правдива , чи неправдива інформація. Після проведення обчислень отримуємо дані які будуть наведені у табл. 2.

Таблиця 2

**Данні розрахунків класифікатора за запропонованим методом з використанням
теореми Байєса для прийняття рішення**

Варіант	1	2	3	4	5	6
Апіорна ймовірність правдивої інформації	0,8	0,7	0,6	0,4	0,3	0,2
Апіорна ймовірність неправдивої інформації	0,2	0,3	0,4	0,6	0,7	0,8
Коефіцієнт прийняття рішення	0,85	0,86	0,87	0,88	0,89	0,9

Аналіз отриманих результатів показує, що зі зменшенням апіорної ймовірності правдивої інформації зростає нижня границя прийняття рішення. Тобто якщо апіорна ймовірність висока тоді коефіцієнт прийняття рішення менший чим коефіцієнт прийняття рішення при низькій апіорній ймовірності. Що цілком підтверджує роботу здатність запропонованої методики.

Графічне представлення результатів моделювання надано на рис. 1 та рис.2.

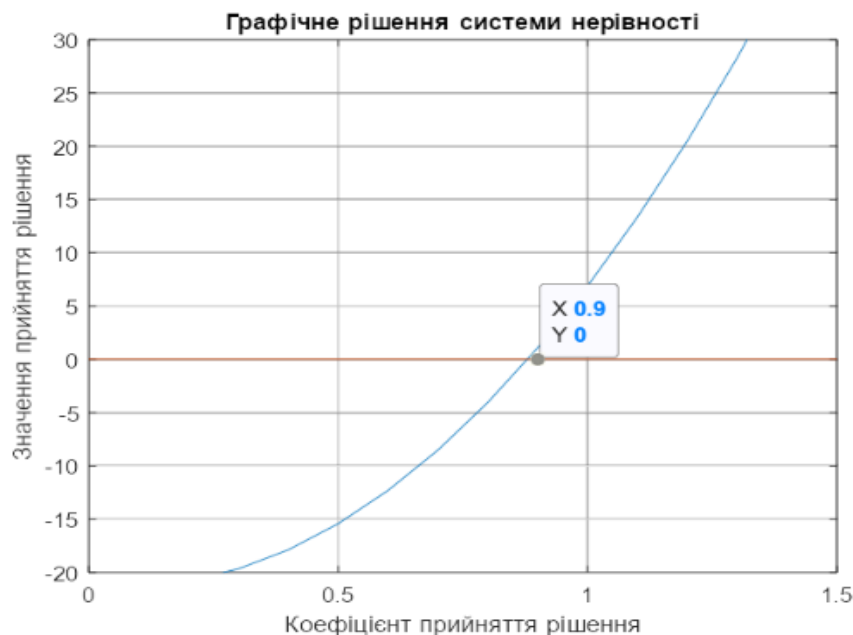


Рис.1. Графік функції ухвалення рішення про визначення правдивої інформації зі застосуванням теореми Байєса (Bayes) для варіанту 6

Аналіз графіка показав, що приблизно після 0,8 коефіцієнт прийняття рішення стає позитивним, що свідчить про те що інформація є правдивою. Графічне рішення

нерівності за виразом (6) цілком співпадає з даними математичних розрахунків, що є додатковим підтвердженням адекватності запропонованого методу розробки класифікатора зі застосуванням теореми Байєса (Bayes) для ухвалення рішення про визначення правдивої інформації.

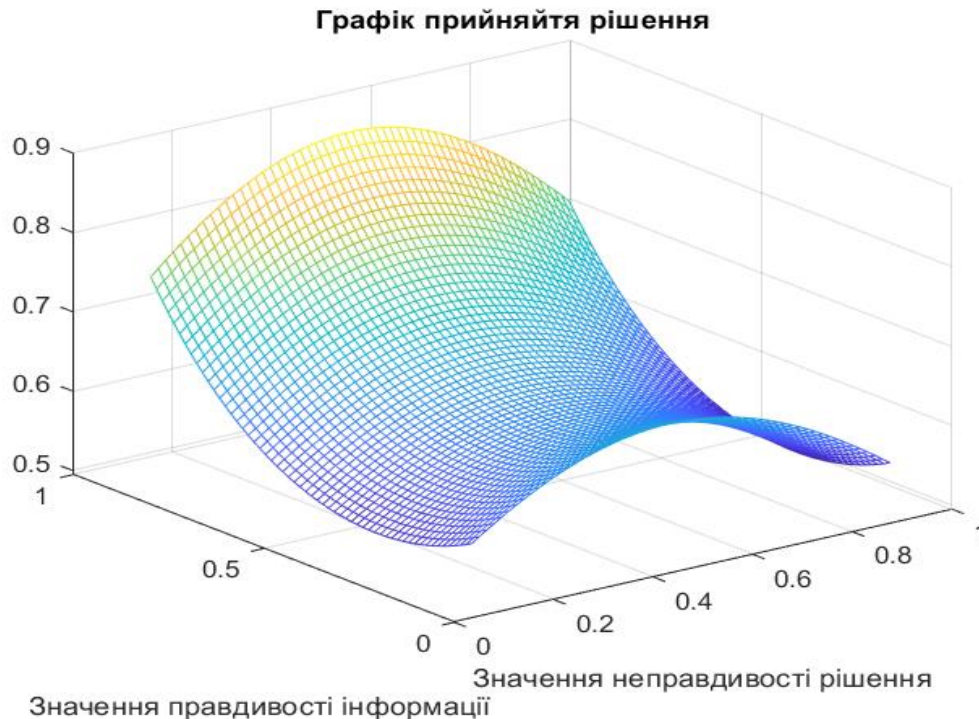


Рис.2. Функції для прийняття рішення, що до правдивості інформації

На рис.2. наведено графік функції прийняття рішення для випадку коли змінюються параметри для правдивості та неправдивості інформації одночасно. Параметри наведені у відносних одиницях.

Отримано графік який показує зростання коефіцієнту прийняття рішення о правдивості інформації при зростанні значення правдивості інформації. Що є підтвердженням адекватності запропонованого методу розробки класифікатора зі застосуванням теореми Байєса (Bayes) для ухвалення рішення про визначення правдивої інформації.

ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

Проведено аналіз існуючих методів класифікації інформації. Сформулювати ознаки інформаційної епохи. В контексті визначення правдивості інформації.

На базі основних ознак інформаційної епохи розроблено метод створення класифікатора для вирішення завдань визначення правдивості інформації.

Проведено математичне моделювання з використанням розробленого класифікатора для підтвердження розробленого методу ухвалення рішення о правдивості інформації за допомогою теореми Байєса (Bayes). Результати розрахунків показали залежність даних розробленого класифікатора від апріорної ймовірності правдивості та неправдивості інформації. Якщо апріорна ймовірність правдивості інформації зменшується тоді коефіцієнт для порівняння про прийняття рішення

збільшується, що доводить адекватність розробленого методу. Але розроблений байєсовський класифікатор ґрунтується на тому, що апіорні ймовірності гіпотез - відомі.

Напрямок подальших досліджень є розробка або удосконалення методів та алгоритмів визначення апіорні ймовірності гіпотез.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

- 1 Наконечний, В., Лаптев, О., Погасій, С., Лазаренко, С., Мартинюк, Г. (2021). Відбір джерел з неправдивою інформацією методом бджолоїної колонії. *Наукоємні технології. Інформаційні технології, кібербезпека*, 52(4), 330-337.
- 2 Zamrii, I., Sobchuk, V., Laptiev, O., Savchenko, V., Shkapa, V., Kovalenko, V., Kotok, V. (2022). Fractal Functions and Their Application to Source Data Coding. *ARPN Journal of Engineering and Applied Sciences*, 17(4), 424 – 435.
- 3 Лаптева, Т. (2021). Алгоритм визначення міри існування недостовірної інформації в умовах інформаційного протиборства. *Кібербезпека: освіта, наука, техніка*, 2(14), 15-25. <https://doi.org/10.28925/2663-4023.2021.14.1525>.
- 4 Наконечний, В. С., Барабаш, О. В., Лаптева, Т. О., Міщенко, А. В. (2022). Удосконалення методу виявлення та кластеризації джерел неправдивої інформації. *Наукоємні технології. Інформаційні технології, кібербезпека*, 54(4), 105 - 111 <https://doi.org/10.18372/2310-5461.54.16747>
- 5 Жигалкевич, Ж.М. (2014). Кластери взаємодіючих підприємств та їх класифікація. *Вісник ОНУ імені І.І. Мечникова*, 19(2/3), 98-101.
- 6 Лаптева, Т.О., Лукова-Чуйко, Н.В., Собчук, А.В. (2022). Дослідження основних загроз і оцінка безпеки інформаційних систем. Математика. Інформаційні технології. Освіта. У *XI Міжнар. наук.–практ. конф., 3–5 червня 2022 р. Луцьк–Світязь: СХУ імені Лесі Українки* (с. 101-103).
- 7 Рябий, М. О., Хатян, О. А., Багацький, С. П. (2015). Модель виявлення PR-впливу через публікації в інтернет ЗМІ. *Інформаційна безпека*, 21(2), 131-139.
- 8 Поліщук, Ю. Я., Гнатюк, С. О., Сейлона, П. А. (2015). Мас медіа як канал маніпулятивного впливу на суспільство. *Інформаційна безпека*, 21(3), 301-308.
- 9 Theocharis, V., Lowe, W., W. van Deth, J., Garcla-Albacete, G. (2015). Using Twitter to mobilize protest action: Online mobilization patterns and action repertoires in the Occupy Wall Street, Indignados, and Aganaktismenoi movements. *Information, Communication & Society*, 18, 202-220.
- 10 Butko, T., Prokhorchenko, A., Muzykin, M. (2016). An improved method of determining the schemes of locomotive circulation with regard to the technological peculiarities of railcar traffic. *Eastern-European Journal of Enterprise Technologies*, 5(3(83)), 47–55. <https://doi.org/10.15587/1729-4061.2016.80471>.
- 11 Молодецька, К. В. (2016). Підхід до виявлення організаційних ознак інформаційних операцій у соціальних інтернет-сервісах. Пріоритетні напрямки розвитку телекомунікаційних систем та мереж спеціального призначення. Застосування підрозділів, комплексів, засобів зв'язку та автоматизації в АТО : збірн. матер. IX наук.-практ. конф., 25 листоп. 2016 р. (с. 130-131). ВІПІ.
- 12 Faraz, A. (2016). A comparison of text Categorization methods. *International Journal on Natural Language Computing*, 5(1), 31 -44.
- 13 Лаптев, О.А., Бабенко, Р.В., Правдивий, А.М., Зозуля, С.А., Стефурак, О.Р. (2020). Удосконалена методика вибору послідовності пріоритетів обслуговування потоків інформації. *Науково-практичний журнал «Зв'язок»*, 4(146), 27 – 31.
- 14 Svynchuk, O., Barabash, O., Nikodem, J., Kochan, R., Laptiev, O. (2021). Image compression using fractal functions. *Fractal and Fractional*, 5(2), 1-14. <https://doi.org/10.3390/fractalfract5020031> - 14 Apr 2021
- 15 Zamrii, I., Sobchuk, V., Laptiev, O., Savchenko, V., Shkapa, V., Kovalenko, V., Kotok, V. (2022). Fractal Functions and Their Application to Source Data Coding. *ARPN Journal of Engineering and Applied Sciences*, 17(4), 424 – 435
- 16 Yevseiev, S., Ponomarenko, V., Laptiev, O., Milov, O. (2021). Synergy of building cybersecurity systems. *Publisher PC TECHNOLOGY CENTER*. <https://doi.org/10.15587/978-617-7319-31-2>

**Nataliya Viktorivna Lukova-Chuiko**

Doctor of technical sciences, professor,
Head of the department of cyber security and information protection
Faculty of Information Technologies
Taras Shevchenko National University of Kyiv, Kyiv, Ukraine
ORCID ID: 0000-0003-3224-4061
lukova@knu.ua

Tetiana Laptieva

PhD student of the department of cyber security and information protection
Faculty of Information Technologies
Taras Shevchenko National University of Kyiv, Kyiv, Ukraine
ORCID ID: 0000-0002-5223-9078
tetiana1986@ukr.net

THE METHOD OF DEVELOPING A CLASSIFIER USING THE BAYES THEOREM FOR MAKING A DECISION ON THE DETERMINATION OF TRUE INFORMATION

Abstract. The range of application of cluster analysis is very wide: it is used in archeology, medicine, psychology, biology, public administration, regional economy, marketing, sociology and other disciplines. Each discipline has its own requirements for primary data and rules for forming groups. Obviously, there will be different methodological approaches to market segmentation, the purpose of which is to identify groups of objects that are similar in terms of features and properties and to the formation of clusters that unite to strengthen their competitive advantages. Thus, when processing information in the information space, the methodology is usually aimed at building a mathematical model of cluster analysis of the object or phenomenon under study, and even obtaining an answer to the question: "Is the information true or not." Detecting false information in the digital world is an important task in overcoming the widespread spread of rumors and prejudices.

The paper analyzes the existing methods of information classification in the information age. Formulate the signs of the information age, in the context of determining the veracity of information. Based on the main features of the information age, a method of creating a classifier has been developed to solve the problems of determining the veracity of information.

Mathematical modeling was carried out using the developed classifier to confirm the developed method of decision-making about the veracity of information using the Bayes theorem. The obtained results proved the efficiency of the proposed method of developing a classifier for which, when applying the Bayes theorem for decision-making, it is possible to determine the veracity of information.

But the developed Bayesian classifier is based on the fact that the a priori probabilities of the hypotheses are known. Therefore, the direction of further research is the development or improvement of methods and algorithms for determining the a priori probability of hypotheses.

Keywords: classifier; clustering; space of images; Bayes theorem; truthfulness of information, a priori probability.

REFERENCES

- 1 Nakonechnyi, V., Laptiev, O., Pohasii, S., Lazarenko, S., Martyniuk, H. (2021). Vidbir dzherel z nepravdyvoiu informatsiuiu metodom bdzholnoi kolonii. Naukoiemni tekhnolohii. Informatsiini tekhnolohii, kiberbezpeka, 52(4), 330-337.
- 2 Zamrii, I., Sobchuk, V., Laptiev, O., Savchenko, V., Shkapa, V., Kovalenko, V., Kotok, V. (2022). Fractal Functions and Their Application to Source Data Coding. ARPN Journal of Engineering and Applied Sciences, 17(4), 424 – 435.
- 3 Laptieva, T. (2021). Alhorytm vyznachennia miry isnuvannia nedostovirnoi informatsii v umovakh informatsiinoho protyborstva. Kiberbezpeka: osvita, nauka, tekhnika, 2(14), 15-25. <https://doi.org/10.28925/2663-4023.2021.14.1525>.



- 4 Nakonechnyi, V. S., Barabash, O. V., Laptieva, T. O., Mishchenko, A. V. (2022). Udoskonalennia metodu vyiavlennia ta klasterizatsii dzherel nepravdyvoi informatsii. Naukoiemni tekhnologii. Informatsiini tekhnologii, kiberbezpeka, 54(4),105 - 111 <https://doi.org/10.18372/2310-5461.54.16747>
- 5 Zhyhalkevych, Zh.M. (2014). Klastery vzaiemodiiuchykh pidprijemstv ta yikh klasyfikatsiia. Visnyk ONU imeni I.I. Mechnykova, 19(2/3), 98-101.
- 6 Laptieva, T.O., Lukova-Chuiko, N.V., Sobchuk, A.V. (2022). Doslidzhennia osnovnykh zahroz i otsinka bezpeky informatsiinykh system. Matematyka. Informatsiini tekhnologii. Osvita. U KhI Mizhnar. nauk.–prakt. konf., 3–5 chervnia 2022 r. Luts'k–Svitiaz: SNU imeni Lesi Ukrainky (s. 101-103).
- 7 Riabyi, M. O., Khatian, O. A., Bahatskyi, C. P. (2015). Model vyiavlennia PR-vplyvu cherez publikatsii v internet ZMI. Informatsiina bezpeka, 21(2), 131-139.
- 8 Polishchuk, Yu. Ya., Hnatiuk, S. O., Seilona, P. A. (2015). Mac media yak kanal manipuliatyvnoho vplyvu na suspilstvo. Informatsiina bezpeka, 21(3), 301-308.
- 9 Theocharis, V., Lowe, W., W. van Deth, J., Garcla-Albacete, G. (2015). Using Twitter to mobilize protest action: Online mobilization patterns and action repertoires in the Occupy Wall Street, Indignados, and Aganaktismenoi movements. Information, Communication & Society, 18, 202-220.
- 10 Butko, T., Prokhorchenko, A., Muzykin, M. (2016). An improved method of determining the schemes of locomotive circulation with regard to the technological peculiarities of railcar traffic. Eastern-European Journal of Enterprise Technologies, 5(3(83)), 47–55. <https://doi.org/10.15587/1729-4061.2016.80471>.
- 11 Molodetska, K. V. (2016). Pidkhid do vyiavlennia orhanizatsiinykh oznak informatsiinykh operatsii u sotsialnykh internet-servisakh. Priorytetni napriamky rozvytku telekomunikatsiinykh system ta merezh spetsialnoho pryznachennia. Zastosuvannia pidrozdiliv, kompleksiv, zasobiv zviazku ta avtomatyzatsii v ATO: zbirn. mater. IX nauk.-prakt. konf., 25 lystop. 2016 r. (c. 130-131). VITI.
- 12 Faraz, A. (2016). A comparison of text Categorization methods. International Journal on Natural Language Computing, 5(1), 31 -44.
- 13 Laptiev, O.A., Babenko, R.V., Pravdyvyi, A.M., Zozulia, S.A., Stefurak, O.R. (2020). Udoskonalena metodyka vyboru poslidovnosti priorytetiv obsluhovannia potokiv informatsii. Naukovo-praktychnyi zhurnal «Zviazok», 4(146), 27 – 31.
- 14 Svynchuk, O., Barabash, O., Nikodem, J., Kochan, R., Laptiev, O. (2021). Image compression using fractal functions. Fractal and Fractional, 5(2), 1-14. <https://doi.org/10.3390/fractalfract5020031> - 14 Apr 2021
- 15 Zamrii, I., Sobchuk, V., Laptiev, O., Savchenko, V., Shkapa, V., Kovalenko, V., Kotok, V. (2022). Fractal Functions and Their Application to Source Data Coding. ARPN Journal of Engineering and Applied Sciences, 17(4), 424 – 435
- 16 Yevseiev, S., Ponomarenko, V., Laptiev, O., Milov, O. (2021). Synergy of building cybersecurity systems. Publisher PC TECHNOLOGY CENTER. <https://doi.org/10.15587/978-617-7319-31-2>

