ISSN 2524-2601 (Online)
ISSN 2304-6201 (Print)

Міністерство освіти і науки України
Харківський національний університет імені В. Н. Каразіна

ВІСНИК

Харківського національного університету імені В.Н. Каразіна

Серія
«Математичне моделювання.
Інформаційні технології.
Автоматизовані системи управління»

Випуск 61

Серія заснована 2003 р.

BULLETIN

of V.N. Karazin Kharkiv National University

Series
«Mathematical Modeling.
Information Technology.
Automated Control Systems»

Issue 61

First published in 2003

Харків
2024

Засновник журналу Харківський національний університет імені В. Н. Каразіна, Харків, Україна. Рік заснування 2003. Періодичність: 4 випуски на рік. https://periodicals.karazin.ua/mia

Статті містять дослідження у галузі математичного моделювання та обчислювальних методів, інформаційних технологій, захисту інформації. Висвітлюються нові математичні методи дослідження та керування фізичними, технічними та інформаційними процесами, дослідження з програмування та комп'ютерного моделювання в наукоємних технологіях.

Для викладачів, наукових працівників, аспірантів, працюючих у відповідних або суміжних напрямках.

Наказом Міністерства освіти і науки України від 17.03.2020 № 409 наукове фахове періодичне видання Вісник Харківського національного університету імені В.Н. Каразіна серія «Математичне моделювання. Інформаційні технології. Автоматизовані системи управління» включено до Категорії «Б» Переліку наукових фахових видань України за наступними спеціальностями: 113 – Прикладна математика;122 – Комп'ютерні науки та інформаційні технології; 123 – Комп'ютерна інженерія; 125 – Кібербезпека.

Затверджено до друку рішенням Вченої ради Харківського національного університету імені В. Н. Каразіна (протокол № 10 від 27.05.2024 р.)

Адреса редакційної колегії: 61022, м. Харків, майдан Свободи, 6, ХНУ імені В. Н. Каразіна, к. 534. Тел. +380 (57) 705-42-81, Email: journal-mia@karazin.ua.

Мова публікації: українська, англійська.

Статті пройшли внутрішнє та зовнішнє рецензування.

Ідентифікатор медіа у Реєстрі суб'єктів у сфері медіа: R30-04456
(Рішення № 1538 від 09.05.2024 р Національної ради України з питань телебачення і радіомовлення. Протокол № 15)

The founder of the Journal is V. N. Karazin Kharkiv National University, Kharkiv, Ukraine.
Year of foundation 2003. The journal is published four times a year.
https://periodicals.karazin.ua/mia

The articles are present research in the field of mathematical modeling and computing methods, information technologies, information security. New mathematical methods of research and management of physical, technical and information processes, research on programming and computer modeling in science-intensive technologies are covered.

For teachers, researchers, graduate students working in relevant or related fields.

By the order of the Ministry of Education and Science of Ukraine from 17.03.2020 № 409 scientific professional periodical Bulletin of V.N. Karazin Kharkiv National University series "Mathematical modeling. Information Technologies. Automated control systems" is included in Category "B" of the List of scientific professional publications of Ukraine in the following specialties: 113 – Applied Mathematics, 122 – Computer Science and Information Technology; 123 – Computer engineering; 125 – Cybersecurity.

Approved for publication by the decision of the Academic Council of V.N. Karazin Kharkiv National University (Minutes № 10 of 27.05.2024).

Editorial Address: 61022, Kharkiv, Svobodi sq., 6, V.N. Karazin Kharkiv National University, r. 534. Phone. +380 (57) 705-42-81, Email: journal-mia@karazin.ua.

Language of publication: Ukrainian, English.

The articles pass internal and external review.

Media identifier in the Register of the field of Media Entities: R30-04456
(Decision № 1538 dated May 9, 2024 of the National Council of Television and Radio Broadcasting of Ukraine, Protocol № 15)

ЗМІСТ

CONTENTS

DOI: https://doi.org/10.26565/2304-6201-2024-61-01

УДК (UDC) 004.08

Бакуменко Ніна Станіславівна *к.т.н., доцент, доцент кафедри теоретичної та прикладної системотехніки, Харківський національний університет імені В.Н. Каразіна, майдан Свободи, 4, Харків-22, Україна, 61022; e-mail: n.bakumenko@karazin.ua;* **https://orcid.org/0000-0003-3496-7167**

Толстолузька Олена Геннадіївна *д.т.н., с.н.с., професор кафедри теоретичної та прикладної системотехніки; Харківський національний університет імені В.Н. Каразіна, майдан Свободи, 4, м. Харків-22, Україна, 61022 e-mail: elena.tolstoluzka@karazin.ua;* **https://orcid.org/0000-0003-1241-7906**

Ясінський Ярослав Андрійович *студент; Харківський національний університет імені В.Н. Каразіна, майдан Свободи, 4, Харків-22, Україна, 61022; e-mail: yar.yasinskyi@gmail.com;* **https://orcid.org/0009-0008-0460-5687**

Аналіз алгоритмів кластеризації для надання рекомендацій товарів

Актуальність. У сучасному світі, насиченому широким спектром товарів та послуг, питання надання персоналізованих рекомендацій для вибору стає актуальним завданням для багатьох сфер, зокрема електронної комерції та онлайн-платформ. Рекомендаційні системи, що працюють на основі пошукових алгоритмів та алгоритмів кластеризації, мають потенціал для значного покращення користувацького досвіду, пропонуючи релевантні та персоналізовані пропозиції товарів. Одними з ключових переваг використання алгоритмів кластеризації для рекомендаційних систем є можливість прогнозувати схожість елементів в залежності від відповідності до певної характеристики, завдяки чому можливо реалізувати ефективний пошук товарів за характеристиками. Внаслідок чого з'являється можливість є сегментувати базу користувачів на окремі підгрупи, що можуть представляти різні сегменти ринку, групи за вподобаннями, цільову аудиторію певних товарів. Виявлення проблем і недоліків таких систем дозволяє вдосконалювати алгоритми, що призводить до більш точних прогнозів і та збільшення продажів компаній.
Мета. Мета даної статті полягає в аналізі ефективності використання методів кластерного аналізу в задачах формування рекомендацій.
Методи дослідження. Порівняльний аналіз, експеримент.
Результати. Проведено аналіз ефективності алгоритмів кластеризації різних типів (k -means++, Mean Shift та HDBSCAN) для надання рекомендацій товарів на основі оцінювання відповідності запиту користувача у відсотковому відношенні, використання оперативної пам'яті, та час виконання запиту. Серед розглянутих найкращі характеристики показав алгоритм k-means++.
Висновки. Проведений аналіз підтверджує ефективність використання методів кластерного аналізу в рекомендаційних системах. Виявлення проблем і недоліків таких систем дозволяє вдосконалювати алгоритми, що призводить до більш точних прогнозів і та збільшення продажів компаній.

Ключові слова: *алгоритм кластеризації, рекомендаційна система, k-means++, HDBSCAN, Mean Shift.*

Як цитувати: Бакуменко Н. С., Толстолузька О. Г., Ясінський Я. А. Аналіз алгоритмів кластеризації для надання рекомендацій товарів. *Вісник Харківського національного університету імені В.Н.Каразіна, серія Математичне моделювання. Інформаційні технології. Автоматизовані системи управління*. 2024. вип. 61. С.6-13. https://doi.org/10.26565/2304-6201-2024-61-01

How to quote: N.S. Bakumenko, O. G. Tolstoluzka, Y. A. Yasinskyi, "Analysis of clustering algorithms for providing product recommendations", *Bulletin of V. N. Karazin Kharkiv National University, series Mathematical modeling. Information Technology. Automated control systems,* vol. 61, pp.6-13, 2024. [In Ukrainian]. https://doi.org/10.26565/2304-6201-2024-61-01

1 Вступ

В сучасному світі електронної комерції успішне надання рекомендацій товарів є важливою складовою для підвищення якості торгівельних сервісів, ефективності реклами та збільшення прибутків компаній. Ефективна система рекомендацій може сприяти підвищенню конкурентоспроможності платформи, збільшуючи продажі завдяки персоналізованим пропозиціям. Системи рекомендацій, що пропонують товари відповідно до індивідуальних уподобань користувачів, значно покращують досвід покупців, стимулюючи їх до повторних покупок. Одним із методів, який використовується в таких системах, є кластеризація даних.

Кластеризація – це процес розподілу набору об'єктів на групи (кластери) таким чином, щоб об'єкти в одному кластері були більш схожі один на одного, ніж на об'єкти в інших кластерах. Застосування алгоритмів кластеризації в системах рекомендацій дозволяє групувати користувачів або товари на основі їх характеристик та поведінки, що в свою чергу сприяє наданню більш релевантних рекомендацій.

У даній роботі статті досліджується ефективність різних алгоритмів кластеризації для надання рекомендацій товарів. Основною метою є аналіз та порівняння цих алгоритмів з точки зору точності рекомендацій та їх відповідності очікуванням користувачів. Для досягнення цієї мети було створено спеціальний набір даних з реальної бази існуючих товарів, реалізовано методи надання рекомендацій на основі кластеризації, та проведено оцінка ефективності роботи цих методів.

2 Використання алгоритмів кластеризації для надання рекомендацій

Кластеризація є методом, який часто використовується для рекомендаційних систем, оскільки вона дозволяє ідентифікувати групи користувачів зі схожими смаками. Це сприяє більш цілеспрямованим рекомендаціям, використовуючи переваги отримання інформації вже визначених вподобань клієнтів [1, 2]. В роботі [3] було показано, що кластеризація може допомогти подолати такі проблеми, як розрідженість даних і масштабованість, які часто виникають при формуванні рекомендацій, і, таким чином, сприяти побудові більш точних і різноманітних рекомендацій.

2.1 Кластеризація на основі центроїдів

Розглянемо групу алгоритмів кластеризації, робота яких заснована на визначенні центрів кластерів. Принцип роботи алгоритму побудований на використанні групування подібних точок властивостей товарів у кластери шляхом встановлення репрезентативних точок, які є центроїдами. Відомим алгоритмом на основі центроїда є k-means, який працює шляхом виділення подібних точок даних, наприклад, користувачів або елементів у кластери, що представлені центроїдами. Головним недоліком цього алгоритму у персоналізованих системах рекомендацій є чутливість до стартового вибору центроїдів. Якщо початкові центроїди не вибрані ретельно, алгоритм може сходитися до локального оптимуму, що призведе до неоптимальної кластеризації [4, 5].

Існує покращена версія алгоритму, що має назву k-means++ – це варіант k-means, що усуває його основний недолік, тобто чутливість до початкових умов. Центр першого кластеру обрається випадково, а центр наступного кластера обирається як найбільш віддалений від попереднього. Це усуває основний недолік алгоритму, пов’язаний з чутливістю до вибору початкових даних [6].

2.2 Ієрархічна кластеризація

Ієрархічна кластеризація базується на підході аналізу елементів, що більш пов’язані з елементами поряд, а ніж з тими, що розташовані далі. Цей тип алгоритмів створює деревоподібні структури з метою генерації кластерів. Основним підходом для створення кластера є підбір елементів зі структури, за відстанню, тоді створений фрагмент даних характеризується максимальною відстанню, для групування частин кластера.

Цікавим прикладом є алгоритм HDBSCAN, що ієрархічним алгоритм кластеризації, який особливо підтвердив свою ефективність у використанні для персоналізованих систем рекомендацій. Цей алгоритм кластеризації базується на щільності розподілу, що означає, що він групує точки датасету разом на основі їх щільності в просторі даних. Це робить HDBSCAN більш стійким до викидів і шуму, ніж інші алгоритми кластеризації [7]. Додатковою перевагою визначення кількості кластерів в процесі роботи алгоритму, н відміну від k-means, і можливість ідентифікувати кластери різної форми.

Задля використання HDBSCAN для персоналізованих систем рекомендацій, першим кроком є об’єднання користувачів у групи зі схожими перевагами, що можна зробити шляхом кластеризації користувачів на основі їхніх попередніх оцінок елементів, таких як фільми, книги чи продукти. Після об’єднання користувачів у кластери наступним кроком є створення персоналізованих рекомендацій для кожного користувача. Це можна зробити, порекомендувавши елементи, популярні серед користувачів у визначеному кластері.

2.3 Кластеризація на основі щільності

Кластеризація на базі щільності використовує ідею ідентифікації груп елементів в даних через припущення, що кластер у просторі даних є безперервною областю високої щільності, відокремленої від інших таких кластерів суміжними областями низької точки щільності. Точки даних у роздільних областях із низькою щільністю точок зазвичай вважаються шумом чи викидами.

Розглянемо цей тип кластеризації на прикладі алгоритму Mean shift. Перший крок передбачає оцінку базової функції щільності ймовірності розподілу точок даних. Зазвичай це робиться за допомогою оцінки щільності ядра, де кожна точка даних представлена функцією ядра з центром у цій точці. Функція ядра визначає вагу, призначену кожній точці даних у процесі оцінки щільності. Наступним кроком алгоритм ітеративно переміщує точки даних до областей з вищою щільністю на основі векторів середнього зсуву, обчислених як зважене середнє значення відмінностей між кожною точкою та її сусідами. Ітерації тривають до конвергенції, що вказується векторами мінімального середнього зсуву. Кінцеве розташування кожної точки даних позначає центр кластера, що дозволяє призначити найближчий центр і ідентифікувати окремі кластери в даних [8].

На відміну від добре відомого підходу кластеризації k-means, Mean shift не потребує припущень щодо кількості кластерів і форми розподілу, але його продуктивність залежить від вибору параметрів масштабу. Пропускна здатність є єдиним параметром для налаштування, тому для одновимірного випадку це відносно проста процедура, але в багатовимірному випадку це може викликати певні складнощі [9].

Для розгляду в контексті використання в рекомендаційних системах були обрані три алгоритми, такі як k-means++, HDBSCAN та Mean Shift, кожен з яких відноситься різних типів кластеризації.

3 Тестовий набір даних

Для тестування роботи алгоритмів було створено датасет з реальними даними про мікрохвильові печі з українських відкритих джерел, з переліком характеристик цифрових та категоріальних. Для тестування надання персоналізованих рекомендацій було обрано об'єм датасету розміром 100 елементів. Дані для датасету про існуючі товари з отримані з агрегаторів товарів у листопаді 2023 року.

Створений датасет містить 9 типів характеристик, що описують товар, окрім id_product, що є ідентифікаційним номером, та не використовується в обчисленнях. Категоріальними даними є колір, виробник та назва виробу. Характеристики ширини, глибини та висоти, містять значення з плаваючою точкою (рис. 1). Ціна, потужність, та максимальна споживана потужність – цілі числа.

	A	B	C	D	E	F	G	H	I	J
1	id_product	manufacturer	name	cost	color	width	height	depth	power	max_consumption
2	1	Gorenje	MO17E1B	2689	black	45.5	26.1	35.3	700	1150
3	2	ERGO	Y35MW	2313	white	44.6	24.3	33.8	700	1100
4	3	Edler	ED-2067W	1961	white	44.4	24.1	35.8	700	1150
5	4	MILANO	MW-4001W	1999	white	45	25	37	700	1100
6	5	Grunhelm	20MX701-W	1899	white	44.8	24.5	32.9	700	1000
7	6	ERGO	EM-2040	2070	black	44	25.9	34.3	700	1050
8	7	Edler	ED-2079W	2059	white	45.1	25.9	33.5	700	1150
9	8	Mirta	Elegance MW-2510W	1900	white	44.6	24.3	33.2	800	1100
10	9	Hansa	AMGF17M2BH	2299	black	45.2	26.2	31.5	600	1000
11	10	Liberton	LMW-2077M	2165	black	44	35.5	25.9	700	1100

Рисунок 1. Приклад перших 10 записів сформованого набору даних

4 Опис обраних технологій

Програмна реалізація була виконана на мові Python з використанням функцій бібліотеки Scikit-learn. Пакет надає набір інструментів для таких завдань, як розробка комп'ютерних моделей, калькуляція метрик для оцінювання ефективності, а також різноманітні доповнення, що наприклад включають функції попередньої обробки інформації датасету [10]. Також було використано бібліотеку Pandas, яка надає інтсрументи для різноманітної маніпуляції даними, зокрема імпорту даних з баз даних, електронних таблиць, файлів формату CSV, та інших [11].

Для оцінки використання пам'яті програмою був використаний профілізатор пам'яті, основне призначення якого виявлення витоків пам'яті та покращення використання пам'яті у програмах на Python. Профілізатор пам'яті аналізує ефективність використання місця в коді та характеристики використовуваних пакетів, пропонуючи ідеї для оптимізації використання пам'яті. Пакет memory_profiler перевіряє використання пам'яті інтерпретатором на кожному рядку. Стовпчик інкрементів дозволяє виявити місця в коді, де виділяються великі обсяги пам'яті. Це особливо важливо при роботі з масивами [12].

Розрахунки проводилися на базі процесору Intel(R) Core(TM) i7-7700HQ CPU, з частотою 2,80 ГГц. Вбудована оперативна пам'ять - 16,0 ГБ.

5 Оцінка ефективності алгоритмів кластеризації у системі рекомендацій

Для реалізації системи рекомендацій на базі кластеризації необхідно дослідити ефективність роботи моделі на різних алгоритмах кластеризації. Для оцінки ефективності надання рекомендацій були використані три параметри: відповідність запиту користувача у відсотковому відношенні, використання оперативної пам'яті, та час виконання запиту.

Перший кроком проводилася кластеризація на основі введеного масиву характеристик. Як результат, були отримані рекомендації на основі відповідного запиту кластеру.

Слід зазначити, запит від користувача можна розглядати як ідеальний товар, що йому необхідний. Набір даних може не містити обраного елементу. Тоді задачею моделі пошук елементів найбільш подібних обраному, та розрахунок цієї подібності.

Після цього відбувається перевірка вимогам користувача за допомогою розрахунку евклідової відстані для обраного товару та кожного елементу у цьому списку [13].

Для коректної роботи алгоритму k-means++ необхідно було визначити кількість кластерів. Для цього можна використати метод ліктя, що визначає оптимальну кількість кластерів у наборі даних шляхом побудови залежності деякої цільової функції якості кластеризації від кількості кластерів.

На рис. 2 зображено графік залежності внутрішньокластерної дисперсії від кількості кластерів для створеного набору даних. «Точка ліктя» розташована там, де швидкість зниження різко змінюється, та вказує на оптимальну кількість кластерів для використання в алгоритмі кластеризації [14], у даному випадку – визначена кількість кластерів дорівнює 3.

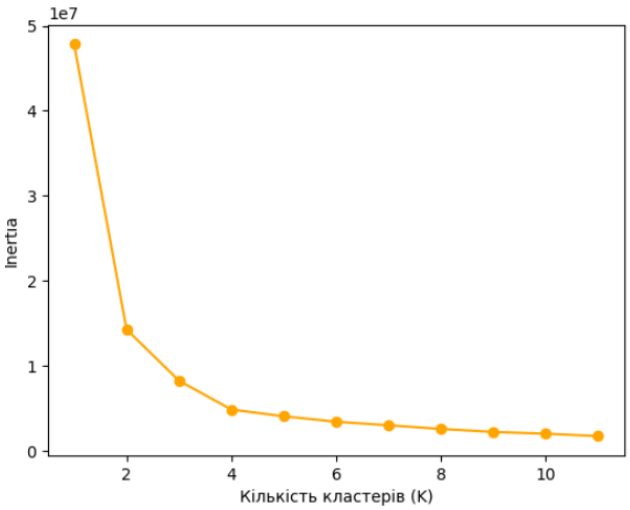


Рисунок 2. Графік залежності внутрішньокластерної дисперсії від кількості кластерів

Для оцінки точності роботи моделі, нам необхідно ввести характеристики наявного товару з набору даних та перевірити, як надаються рекомендації. Якщо евклідова відстань введеного

товару близька до 0, то можна констатувати 100 відсотковий збіг – отже, модель працює правильно. Для порівняння результатів різних алгоритмів необхідно порівняти відстані схожих елементів, що відрізняються за однією чи кількома характеристиками. До прикладу, оберемо першу мікрохвильову піч з характеристиками зазначеними на рис. 1.

Перший продукт у таблиці (M017E1B) має показник евклідової відстані 0,00. Це тому, що продукт ідентичний сам собі. Всі алгоритми змогли визначити наявність цього елементу в наборі даних, отже створена реалізація працює коректно. Інші елементи в таблиці мають оцінку схожості, що перевищує 0,00. Це означає, що продукт схожий на M017E1B, та може бути запропонований користувачу.

Після запуску моделі на основі алгоритму k-means++ модель може визначити який елемент був введений через те, що модель виставила 100 відсотків збігу введеному елементу. Додатково отримано список десяти рекомендацій товарів (рис. 3).

Другий продукт (M017E1W) є схожим на товар, на основі якого виконуються рекомендації, але не таким самим, оскільки в нього відмінне одне значення – колір, тоді різниця в один категоріальний параметр складає 10.07.

Останній елемент у списку рекомендацій k-means ++ (PMW 20711 KB) має значення – 331.45**,** отже має схожість з обраним товаром, на основі якого робляться рекомендації, але може бути значно відмінним.

Результати ієрархічної кластеризації гірші аніж двох методів перед ним. Найбільша відстань складає 796.86.

Рекомендації K-means++			Рекомендації Mean Shift			Рекомендації HDBSCAN		
Id	Product	Similarity	Id	Product	Similarity	Id	Product	Similarity
0	M017E1B	0.00	0	M017E1B	0.00	0	M017E1B	0.00
18	M017E1W	10.07	18	M017E1W	10.07	18	M017E1W	10.07
19	M020E1WH	130.34	13	PMW 20757 HB	100.35	17	PMW 20711 KW	126.33
22	M017E1S	152.00	17	PMW 20711 KW	126.33	19	M020E1WH	130.34
11	MW-4001BR	199.42	19	M020E1WH	130.34	11	MW-4001BR	199.42
20	LMW-2074M	211.43	11	MW-4001BR	199.42	24	MW-MM-20P(WH)	220.47
28	M017E1BH	251.16	14	PMW 20711 KB	331.45	34	AMG20M70GSVH	398.25
29	LMW-2079M	284.72	1	Y35MW	379.32	8	AMGF17M2BH	429.67
31	R200BKW	322.38	15	PMW 20715 KB	379.42	45	MW-4010B	612.05
14	PMW 20711 KB	331.45	12	PMW 20757 HW	382.12	7	Elegance MW-2510W	796.89

Рисунок 3. Списки рекомендованих товарів створених за допомогою різних алгоритмів кластеризації

Для визначення кращого методу для надання персоналізованих рекомендацій було обрано збіжні елементи з кількох списків, отриманих із застосуванням різних алгоритмів кластеризації.

Для порівняння ефективності надання рекомендацій запропонованих методів підраховано середню суму відстаней, максимальну відстань та підсумкову суму відстаней для усіх кластерів.

Результати розрахунків задані у табл. 1:

Таблиця 1. Порівняння результатів надання рекомендацій

Евклід. відстань	K-means++	Mean Shift	HDBSCAN
Максимальна	331.45	382.12	796.89
Середня	210.33	237.5	324.83
Сума	1892.97	2137.57	2923.49

Таблиця містить порівняльний аналіз трьох алгоритмів кластеризації – k-means++, Mean Shift та HDBSCANз використанням евклідової відстані для оцінки точності рекомендацій товарів. Евклідова відстань вимірює схожість між рекомендованими товарами та запитаним товаром.

Для тестування часу було вирішено провести декілька ітерацій запуску моделі, щоб переконатися, що результат не був випадковим. Під час кожної спроби вимірювався час, необхідний для виконання моделі кластеризації на тестовому наборі даних. Підсумковим результатом є середнє значення часу виконання для кожного алгоритму. Час роботи алгоритму наведений у табл. 2:

Таблиця 2. Час роботи алгоритму (сек.)

Спроба	K-means++	Mean Shift	HDBSCAN

1	0.86	1.83	0.32
2	0.73	1.79	0.29
3	0.74	1.79	0.33
4	0.73	1.80	0.32
5	0.79	1.81	0.31
Сер. значення	0.77	1.804	0.314

HDBSCAN демонструє найшвидший час виконання серед трьох алгоритмів кластеризації. Середній час виконання складає лише 0.314 секунди, що робить його найбільш ефективним за цим параметром.

K-means++ займає друге місце за швидкістю виконання із середнім часом 0.77 секунди. Хоча він працює повільніше за HDBSCAN, все ж таки час його виконання залишається досить низьким. Результати між декількома запусками одного алгоритму відрізняються на незначну частину, однак найбільша різниця між ітераціями складає 0.09 сек, що може бути пов'язано з першим запуском моделі.

Mean Shift є найповільнішим алгоритмом серед розглянутих, зі середнім часом виконання 1.804 секунди. Це свідчить про його менш ефективну роботу з точки зору часу.

Порівняння обсягу оперативної пам’яті для різних алгоритмів кластеризації наведено в таблиці 3.

Таблиця 3. Використання оперативної пам'яті (MiB)

Спроба	K-means++	Mean Shift	HDBSCAN
1	1.4	1.5	0.8
2	1.4	1.5	0.9
3	1.5	1.6	0.8
4	1.4	1.6	0.9
5	1.4	1.5	0.8
Сер. значення	1.42	1.54	0.84

За результатами експериментів проведених за допомогою Memory profiler за вимірами обсягу пам'яті можна сказати, що найменше споживає пам'яті ієрархічна кластеризація. HDBSCAN демонструє найменше використання оперативної пам'яті серед трьох алгоритмів кластеризації. Середнє використання пам'яті складає лише 0.84 MiB, що робить його найбільш ефективним з точки зору економії пам'яті. Кластеризація на основі щільності показує найбільші результати по споживанню пам'яті – 1.54 MiB, що може пояснюватися тим, що даний підхід, який зазвичай використовується для розпізнавання зображень. K-means++ займає друге місце за обсягом використання пам'яті із середнім значенням 1.42 MiB. Кластеризація на основі центроїдів вимагає трохи менше пам'яті, проте все одно більше ніж HDBSCAN.

6 Висновки

У даному дослідженні проведено аналіз ефективності різних алгоритмів кластеризації для надання рекомендацій товарів. Було розглянуто три алгоритми: k-means++, Mean Shift та HDBSCAN. Для дослідження було створено спеціальний датасет з характеристиками товарів, на основі якого проводилося тестування.

Результати дослідження показали, що алгоритм k-means ++ надає найбільш точні та відповідні рекомендації порівняно з іншими обраними алгоритмами. За результатами тестування k-means ++ демонструє найменшу середню евклідову відстань між рекомендованими товарами та обраним товаром, що вказує на високу точність рекомендацій. Зокрема, максимальна евклідова відстань для K-means++ склала 331.45, що є нижчим показником у порівнянні з Mean Shift (382.12) та HDBSCAN (796.89).

У той же час, алгоритм HDBSCAN показав найкращі результати за часом виконання запитів, що може бути корисним у випадках, коли швидкість є критично важливою. Середній час виконання для HDBSCAN складав лише 0.314 секунд, тоді як для K-means++ та Mean Shift цей показник був значно вищим.

За обсягом використання оперативної пам'яті алгоритм HDBSCAN також показав найкращі результати, споживаючи в середньому 0.84 MiB, що значно менше ніж Mean Shift (1.54 MiB) та K-means++ (1.42 MiB).

Таким чином, вибір алгоритму кластеризації для системи надання рекомендацій товарів залежить від конкретних вимог до системи. Якщо пріоритетом є точність рекомендацій, K-means++ є найкращим вибором.

Отже, результати дослідження підтверджують доцільність використання кластеризації для покращення систем рекомендацій товарів та дозволяють зробити обґрунтований вибір алгоритму залежно від специфічних вимог до системи.

REFERENCES

1. How Search Engine Personalization Affects Rankings. [Online]. Available: https://marketbrew.ai/how-search-engine-personalization-affects-rankings Accessed on: May 21, 2024.
2. Data Clustering: Intro, Methods, Applications. [Online]. Available: https://encord.com/blog/data-clustering-intro-methods-applications Accessed on: May 22, 2024.
3. J. Das, S. Majumder, K. Mali, “Clustering Techniques to Improve Scalability and Accuracy of Recommender Systems*”, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 29, no. 04, pp.. 621–651, 2021
4. k-means Advantages and Disadvantages: [Online]. Available: https://developers.google.com/machine-learning/clustering/ Accessed on: May 22, 2024.
5. Artley B. Unsupervised Learning: k-means Clustering. Towards Data Science: [Online]. Available: https://towardsdatascience.com/unsupervised-learning-k-means-clustering-27416b95af27 Accessed on: May 20, 2024.
6. D. Arthur,S. Vassilvitskii, "k-means++: the advantages of careful seeding", *in Proc. of the Eighteenth annual ACM-SIAM symposium on Discrete algorithms*, Philadelphia, PA, USA., 2007, pp. 1027–1035.
7. Christopher A. Hierarchical Clustering and Density-Based Spatial Clustering of Applications with Noise (DBSCAN): Online]. Available: https://medium.com/mlearning-ai/hierarchical-clustering-and-density-based-spatial-clustering-of-applications-with-noise-dbscan-b8d903095532 Accessed on: May 10, 2024.
8. J. Sander, "Density-Based Clustering", in *Encyclopedia of Machine Learning*,. C. Sammut, G. I. Webb, Eds. Boston, MA, USA:Springer, 2011, pp. 349-353.
9. Damir Demirović, "An Implementation of the Mean Shift Algorithm", *Image Processing On Line*, no. 9, pp. 251–268, 2019.
10. Scikit-learn User Guide: [Online]. Available: https://scikit-learn.org/stable/user_guide.html Accessed on: May 22, 2024.
11. Pandas documentation: [Online]. Available: https://pandas.pydata.org/ Accessed on: May 24, 2024.
12. Memory-profiler: [Online]. Available: https://pypi.org/project/memory-profiler/ Accessed on: May 24, 2024.
13. Euclidean distance score and similarity. Available: https://stats.stackexchange.com/questions/53068/euclidean-distance-score-and-similarity Accessed on: May 24, 2024.
14. Elbow Method for optimal value of k in k-means? Available: https://www.geeksforgeeks.org/elbow-method-for-optimal-value-of-k-in-kmeans/ Accessed on: May 24, 2024.

Bakumenko Nina *Candidate of Technical Sciences; Associate Professor of theoretical and applied system engineering department;*
V.N. Karazin Kharkiv National University
Svobody Sq 4, Kharkiv, Ukraine, 61022
e-mail: n.bakumenko@karazin.ua;
https://orcid.org/0000-0003-3496-7167

Tolstoluzka Olena *doctor of Technical Sciences; Professor of theoretical and applied system engineering department;*
V.N. Karazin Kharkiv National University
Svobody Sq 4, Kharkiv, Ukraine, 61022
e-mail: elena.tolstoluzka@karazin.ua;
https://orcid.org/0000-0003-1241-7906

Yasinskyi Yaroslav *student;*
V.N. Karazin Kharkiv National University
Svobody Sq 4, Kharkiv, Ukraine, 61022
e-mail: yar.yasinskyi@gmail.com;
https://orcid.org/0009-0008-0460-5687

Analysis of clustering algorithms for product recommendations

Relevance. In today's world, where a wide range of goods and services are available, the task of providing personalized recommendations for selecting the right one is becoming an increasingly important in many areas, including e-commerce and online platforms. Expert recommendation systems powered by search and clustering algorithms have the potential to significantly improve the user experience by offering relevant and personalized product suggestions. One of the key advantages of using clustering algorithms for recommender systems is the ability to predict the similarity of objects based on their compliance with a certain characteristic, which makes it possible to implement an effective search for products by characteristics. As a result, it allows dividing an user base into separate subgroups that can represent different market segments, preference groups, and the target audience of certain products. Identification of problems and shortcomings of such systems helps to improve algorithms, which leads to more accurate forecasts and increased sales.
Objective. The purpose of this article is to analyze the effectiveness of using cluster analysis methods in the tasks of generating recommendations.
Research methods. Comparative analysis, experiment.
Results. The effectiveness of clustering algorithms of different types (k-means++, Mean Shift and HDBSCAN) for providing product recommendations based on the assessment of the percentage of compliance with the user's request, the use of RAM, and the query execution time has been analyzed. The k-means++ algorithm showed the best performance among the tested algorithms.
Conclusions. Our analysis confirms the effectiveness of using cluster analysis methods in recommender systems. Identification of problems and shortcomings of such systems allows improving algorithms, which leads to more accurate forecasts and increased sales of companies.

Keywords: *clustering algorithm, recommender system, k-means++, HDBSCAN, Mean Shift.*

Надійшла у першій редакції 18.02.2024, в останній - 21.04.2024.
The first version has been received on 18.02.2024, the final version - on 21.04.2024.

DOI: https://doi.org/10.26565/2304-6201-2024-61-02

УДК 539.3+519.60

Дегтярьов Кирило Георгійович *к.т.н., науковий співробітник Інститут проблем машинобудування ім. А.М. Підгорного НАН України, , м. Харків, вул. Комунальників, 2/10, 61023 e-mail: kdegt89@gmail.com;* **https://orcid.org/0000-0002-4486-2468**

Колодяжний Андрій Сергійович *аспірант Інститут проблем машинобудування ім. А.М. Підгорного НАН України, м. Харків, вул. Комунальників, 2/10, 61023 e-mail:7ask7@ukr.net;* **https://orcid.org/0000-0008-4026-6715**

Крютченко Денис Володимірович *доктор філософії., молодший науковий співробітник Інститут проблем машинобудування ім. А.М. Підгорного НАН України, , м. Харків, вул. Комунальників, 2/10, 61023 e-mail: wollydenis@gmail.com;* **https://orcid.org/0000-0003-6804-6991**

Усатова Ольга Олександрівна *інженер Інститут проблем машинобудування ім. А.М. Підгорного НАН України, , м. Харків, вул. Комунальників, 2/10, 61023 e-mail: usatova.olia@gmail.com;* **https://orcid.org/0000-0001-1267-2723**

Стрельнікова Олена Олексанрівна *д.т.н., проф., провідний науковий співробітник Інститут проблем машинобудування ім. А.М. Підгорного НАН України м. Харків, вул. Комунальників, 2/10, 61023 e-mail: maxim.sidorov@nure.ua;* **https://orcid.org/0000-0003-0707-7214**

Комп’ютерне моделювання плескань рідини в резервуарах при періодичних навантаженнях

Основною метою роботи є розроблення комп’ютерної методології для стійкості руху рідини в резервуарах та паливних баках під дією періодичних зовнішніх з урахуванням демпфування.

Актуальність. Демпфування відіграє вирішальну роль у забезпеченні стабільності та зменшенні потенційних небезпек при експлуатації резервуарів, частково заповнених рідиною. Відсутність амортизації може призвести до нестабільності руху. У резервуарах з рідиною будь-які порушення руху, такі як раптове прискорення, уповільнення або поворот, можуть спричинити плескання. Без амортизації, плескання можуть навіть посилюватися, потенційно призводячи до неконтрольованих і небезпечних ситуацій, особливо в транспортних засобах або промислових процесах. Демпфування забезпечує контроль над динамікою плескань, забезпечуючи більш плавну та передбачувану поведінку. За допомогою гасіння надмірних коливань інженери можуть гарантувати, що рідина залишається стабільною всередині паливного бака, що зменшує ризик надмірних динамічних навантажень на конструкцію бака або транспортний засіб, який його перевозить. Тому є актуальними дослідження, присвячені вивченню демпфування плескань.

Методи дослідження. Для розв’язання задачі демпфування плескань використані методи інтегральних рівнянь, метод заданих форм та метод граничних елементів.

Результати. Розв’язано спектральну граничну задачу та знайдено частоти та форми власних коливань рідини ходження власних частот та форм коливань рідини в жорсткому Стійкість руху при вертикальних гармонічних навантаженнях визначено за допомогою діаграми Айнса-Стретта. Досліджено комбіновані горизонтальні та вертикальні навантаження, та знайдені зони стійкого та нестійкого руху в залежності від параметрів навантаження. Вивчено вплив демпфування з використанням матриці Релея. Важливість отриманих результатів щодо плескань рідини в жорстких резервуарах полягає в з’ясуванні вирішальної ролі демпфування у забезпеченні стабільності та зменшенні потенційних небезпек щодо стійкості паливних баків ракет-носіїв під час польоту.

Висновки. Розроблено метод визначення змінного за часом рівня вільної поверхні рідини в жорстких оболонках обертання. Спектральна задача з визначення частот та форм коливань рідини в усіченому конічному резервуарі розв’язана шляхом зведення до системи одновимірних інтегральних рівнянь. За допомогою діаграми Айнса-Стретта знайдені зони нестійкості руху рідини при гармонічних вертикальних навантаженнях. З’ясовано вплив демпфування за Релеєм на зростання рівня вільної поверхні. В подальшому передбачається дослідження коливань пружних оболонок обертання з рідиною, з використанням різних композитних матеріалів.

Ключові слова: *вільна поверхня, плескання рідини в резервуарах, системи сингулярних інтегральних рівнянь .метод граничних елементів, демпфування, діаграма Айнса-Стретта.*

Як цитувати: Дегтярьов К.Г., Колодяжний А.С., Крютченко Д.В., Усатова О.О., Стрельнікова О. О. Комп'ютерне моделювання плескань рідини в резервуарах при періодичних навантаженнях. *Вісник Харківського національного університету імені В.Н.Каразіна, серія Математичне моделювання. Інформаційні технології. Автоматизовані системи управління.* 2024. вип. 61. С.14-23. https://doi.org/10.26565/2304-6201-2024-61-02

How to quote: K.G. Degtyarev, A.S. Kolodyazhny, D.V, Kriutchenko, O.O. Usatova and Strelnikova O.O., "Computer modeling of liquid sloshing in tanks under periodic loads" *Bulletin of V.N. Karazin Kharkiv National University, series Mathematical modelling. Information technology. Automated control systems*, vol. 61, pp.14-23, 2024. [In Ukrainian] https://doi.org/10.26565/2304-6201-2024-61-02

1 Вступ

Демпфування відіграє вирішальну роль у забезпеченні стабільності та зменшенні потенційних небезпек при експлуатації резервуарів, частково заповнених рідиною. Амортизація за рахунок демпфування допомагає розсіювати енергію руху плескань рідини. Відсутність амортизації може призвести до нестабільності руху. У резервуарах з рідиною будь-які порушення руху, такі як раптове прискорення, уповільнення або поворот, можуть спричинити плескання. Без амортизації, плескання можуть навіть посилюватися, потенційно призводячи до неконтрольованих і небезпечних ситуацій, особливо в транспортних засобах або промислових процесах. Демпфування забезпечує контроль над динамікою плескань, забезпечуючи більш плавну та передбачувану поведінку. За допомогою гасіння надмірних коливань інженери можуть гарантувати, що рідина залишається стабільною всередині бака, що зменшує ризик надмірних динамічних навантажень на конструкцію бака або транспортний засіб, який його перевозить. Надмірні плескання можуть призвести до явища, яке називається «ефект вільної поверхні», коли поверхня рідини може утворювати значні хвилі або впливати на стінки резервуара зі значною силою. Демпфування допомагає пригнічувати ці хвилі, зменшуючи ймовірність руйнівних ударів і забезпечуючи структурну цілісність бака. Демпфування покращує загальну продуктивність систем, що включають жорсткі баки, наповнені рідиною. Таким чином, демпфування має вирішальне значення при розгляді плескань рідини в жорстких резервуарах через його роль у забезпеченні стабільності, контролю, безпеки, продуктивності обслуговування різних систем і обладнання. Належні механізми амортизації, за допомогою плавучих кришок, перегородок, або їх комбінації, є важливими для оптимізації поведінки наповнених рідиною жорстких резервуарів у широкому діапазоні застосувань.

2 Постановка проблеми та огляд сучасного стану питання

Нові умови використання техніки та нові матеріали призводять до суттєвих змін напружено-деформованого стану, вібраційних характеристик елементів сучасних конструкцій і потребують вдосконалених досліджень міцносних та динамічних характеристик обладнання, яке працює в умовах підвищених силових, температурних факторів, за умови взаємодії з різними заповнювачами. Проблема плескань рідини в резервуарах виникла ще в 60-ті роки минулого століття, коли почалися перші польоти космічних апаратів. Невдале проектування призводило до значних коливань рідини в паливних баків, що в свою чергу, вело до втрати стійкості, сходження з розрахункової траєкторії та навіть повного руйнування ракет-носіїв. Проектування новітніх потужних ракет-носіїв вимагає й нових конструкцій баків, що наразі можуть приймати досить екзотичні форми [1]. Тому вже протягом кількох десятиліть не вщухає інтерес до вивчення проблем стійкості руху в резервуарах та паливних баках [2-4]. Існує велика кількість сучасних ефективних методів для дослідження міцності та коливань елементів сучасного обладнання. Серед них відмітимо метод скінченних елементів (МСЕ) [5], метод скінчених різниць [6], метод граничних елементів (МГЕ) [7], метод скінченних об'ємів [8]. На практиці інженери, що проектують паливні баки для ракет-носіїв, часто використовують різні методи для посилення амортизації, такі як внутрішні перегородки [9], пінопластові вставки [10], повні [11] або часткові [12] покриття вільної поверхні, інноваційні матеріали [13] або системи активного керування [14]. Ці підходи спрямовані на досягнення достатнього демпфування, щоб пом'якшити наслідки плескань, враховуючи такі фактори, як вага, обмеження простору та експлуатаційні вимоги. Використання матриці демпфування Релея є поширеним підходом до включення ефектів демпфування в динамічний аналіз, включаючи плескання в баках ракет-носіїв [15]. Ефекти

демпфування відіграють вирішальну роль при аналізі стійкості руху рідини в резервуарах та паливних баках.

Тому дослідження, присвячене врахуванню демпфування при аналізі стійкості руху рідини в жорстких оболонках обертання, виконано на актуальну тему.

3 Мета дослідження та формулювання задачі

Мета дослідження полягає в розробці комп'ютерної методології для врахування демпфування при аналізі стійкості руху рідини в резервуарах та паливних баках під дією періодичних зовнішніх навантажень.

Розглянуто жорсткі оболонки обертання, частково заповнені рідиною, рис.1. Нехай S_1 - змочена поверхня оболонки, S_0 – вільна поверхня рідини. Припускається, що рідина ідеальна і нестислива. Вважається, що рух рідини є безвихровим. Умова нестисливості має вигляд $\mathrm{div}\mathbf{V} = 0$, де $\mathbf{V}(V_x, V_y, V_z)$ вектор швидкості рідини.

Внаслідок відсутності вихорів існує скалярний потенціал Ф, такий що $\mathbf{V} = \mathrm{grad}\Phi$. Потенціал Ф задовольняє рівнянню Лапласа в області Ω, що зайнята рідиною. Тиск рідини p визначається за формулою [8]

$$\frac{p}{\rho_l} = -\frac{\partial \Phi}{\partial t} - \big(g + a_v(t)\big)z + a_h(t)x + \frac{p_0}{\rho_l}, \tag{3.1}$$

де ρ_l – густина рідини, g – гравітаційна стала, z – вертикальна координата точки всередині рідкого об'єму Ω, $a_h(t)$ та $a_v(t)$ прискорення сили, що змушує, в горизонтальному та вертикальному напрямках.

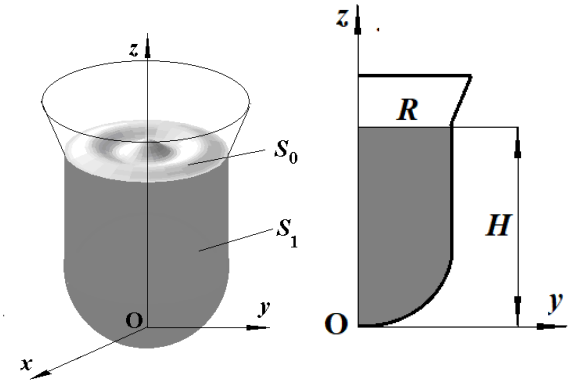


Рис. 3.1. Оболонка обертання з рідиною та її схема

Сформулюємо крайову задачу для рівняння Лапласа у вигляді

$$\nabla^2\Phi = 0, \frac{\partial \Phi}{\partial \mathbf{n}}\Big|_{S_1} = 0,\ \frac{\partial \Phi}{\partial \mathbf{n}}\Big|_{S_0} = \frac{\partial \zeta}{\partial t}, p - p_0|_{S_0} = 0. \tag{3.2}$$

Тут $\mathbf{n}$ – одинична зовнішня нормаль до поверхні, p_0 – атмосферний тиск, $\zeta=\zeta(x, y, t)$ –невідома функція, що описує положення та рух вільної поверхні. Тобто маємо крайову задачу відносно невідомого потенціалу Ф, що пов'язаний кінематичною граничною умовою з функцією $\zeta(x, y, t)$.

4 Метод заданих форм

Зобразимо невідомі функції Ф та ζ в циліндричних координатах у вигляді рядів:

$$\zeta(r, \theta, t) = \sum_{l=0}^{m} \cos(l\theta) \sum_{k=1}^{n} d_{kl}(t)\zeta_k(r), \tag{4.1}$$

$$\Phi(r, \theta, z, t) = \sum_{l=0}^{m} \cos(l\theta) \sum_{k=1}^{n_2} \dot{d}_{kl}(t)\varphi_k(r, z) \tag{4.2}$$

Тут $\varphi_k(r, z)$, $\zeta_k(r)$- базисні функції, між якими на вільній поверхні існує такий зв'язок [16]:

$$\frac{\partial \varphi_k(r,z)}{\partial \mathrm{n}}\Big|_{z=H} = \zeta_k(r). \tag{4.3}$$

При цьому функції $\psi_{kl} = \varphi_{kl}(r, z)\cos(l\,\theta)$ мають задовольняти рівнянню Лапласа. Припускаючи

гармонічний характер зміни коефіцієнтів $d_{kl}(t)$ за часом $d_{kl}(t) = D_{kl}\exp(i\omega_{kl}t)$, отримаємо з (4.3) співвідношення

$$\frac{\partial\psi_{kl}}{\partial\mathbf{n}} = \frac{\omega_{kl}^2}{g}\,\psi_{kl}, \tag{4.4}$$

яке приводить до спектральної крайової задачі відносно ψ_{kl} [16]. Після знаходження розв'язку спектральної задачі отримуємо базисні функції $\varphi_{kl}(r,z)$ та $\zeta_k(r)$ та власні частоти ω_{kl}.

5 Визначальна система диференціальних рівнянь

Вважаємо, що базисні функції $\varphi_{kl}(r,z)$ отримано. Підставимо їх у вирази (4.1) для потенціалу швидкості Ф та (4.2) для висоти підйому вільної поверхні ζ. Далі підставляємо отримані вирази в динамічну умову на S_0. Приходимо до такого співвідношення на вільній поверхні:

$$\sum_{l=0}^{m}\cos(l\theta)\sum_{k=1}^{n}\left[\ddot{d}_{kl}(t) + \omega_{kl}^2\left(1+\frac{a_v(t)}{g}\right)d_{kl}(t)\right]\varphi_{kl}(r,z) + a_h(t)r\cos\theta = 0, z = \zeta. \tag{5.1}$$

Виконав скалярний добуток рівняння (5.1) на функції $\psi_{kl}(k=\overline{1,n}; l=\overline{0,m})$ і використав ортогональність власних форм [17], отримуємо незв'язану систему звичайних диференціальних рівнянь другого порядку

$$\ddot{d}_{k0}(t) + \omega_{k0}^2\left(1+\frac{a_v(t)}{g}\right)d_{k0}(t) = 0, \tag{5.2}$$

$$\ddot{d}_{k1}(t) + \omega_{k1}^2\left(1+\frac{a_v(t)}{g}\right)d_{k1}(t) + a_h(t)F_{k1} = 0,\ F_{k1} = \frac{(r,\varphi_{k1})}{(\varphi_{k1},\varphi_{k1})},$$

$$\ddot{d}_{kl}(t) + \omega_{kl}^2\left(1+\frac{a_v(t)}{g}\right)d_{kl}(t) = 0, k=\overline{1,n}; l=\overline{2,m}.$$

Для однозначного розв'язання системи (5.2) необхідно задати початкові умови, тобто

$$d_{kl}(t) = d_{kl}^0,\ \dot{d}_{kl}(t) = d_{kl}^1,\ k=\overline{1,n},\ l=\overline{0,m} \tag{5.3}$$

Введемо штучне демпфування, як запропоновано в [18], а саме будемо розглядати таку систему диференціальних рівнянь

$$\ddot{d}_{k0}(t) + 2\omega_{k0}c\dot{d}_{k0}(t) + \omega_{k0}^2\left(1+\frac{a_v(t)}{g}\right)d_{k0}(t) = 0, \tag{5.4}$$

$$\ddot{d}_{k1}(t) + 2\omega_{k1}c\dot{d}_{k1}(t) + \omega_{k1}^2\left(1+\frac{a_v(t)}{g}\right)d_{k1}(t) + a_h(t)F_{k1} = 0,\ F_{k1} = \frac{(r,\varphi_{k1})}{(\varphi_{k1},\varphi_{k1})},$$

$$\ddot{d}_{kl}(t) + 2\omega_{kl}c\dot{d}_{kl}(t) + \omega_{kl}^2\left(1+\frac{a_v(t)}{g}\right)d_{kl}(t) = 0, k=\overline{1,n}; l=\overline{2,m}.$$

з початковими умовами (5.3).

Тут c є коефіцієнтом демпфування. В цьому дослідженні обираємо c з інтервалу (0.05,0.01), [19], що відповідає низькому рівню демпфування.

6 Аналіз числових результатів

Як числовий приклад, розглянемо конічну оболонку з рідиною під дією гармонічного навантаження

$$a_x(t) = a_h\cos(\omega_h t), a_z(t) = a_v\cos(\omega_v t). \tag{6.1}$$

Усічена конічна оболонка має такі геометричні параметри: R_1 = 1м, α = π/6. Тут R_1 – радіус вільної поверхні, R_2 – радіус днища, рис. 6.1.

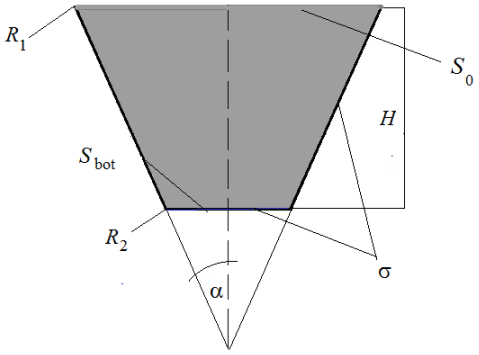


Рис. 6.1. Усічена конічна оболонка з рідиною

Розв'яжемо спектральну крайову задачу та отримаємо форми та частоти плескань цієї оболонки. В таблиці 6.1 наведені частотні параметри $\chi_k^2 = \omega_k^2/g$, за різні значення R_2, обчислені при $l = 0{,}1{,}2$ та $k = 1,$ та здійснено порівняння отриманих результатів з даними роботи [20].

Таблиця 6.1. Частотний параметр плескань рідини в конічній оболонці

	$\chi_k^2 = \omega_k^2/g$			
R_2	0.2	0.6	0.8	0.9
j =0, k =1				
[20]	3.386	3.382	3.139	2.187
МГЕ	3.388	3.392	3.1422	2.200
j =1, k =1				
[20]	1.304	1.254	0.934	0.542
МГЕ	1.305	1.258	0.941	0.564
j=2, k =1				
[20]	2.263	2.255	2.015	1.361
ВЕМ	2.265	2.259	2.028	1.395

Дані таблиці 6.1 демонструють гарну узгодженість результатів, що свідчить про збіжність та вірогідність запропонованого методу. При числових розрахунках обиралось 150 граничних елементів вздовж конічної частини, й по 120 граничних елементів вздовж радіусу вільної поверхні та радіусу днища. Подальше збільшення кількості елементів не привело до суттєвої зміни результатів.

Проведено розрахунки руху вільної поверхні за різні значення параметрів a_h, a_v та $\omega_h,\ \omega_v$. Спочатку розглянемо вертикальні навантаження. Фазові портрети рухів в координатах $(\zeta, \dot{\zeta})$ зображені на рис.6.2.

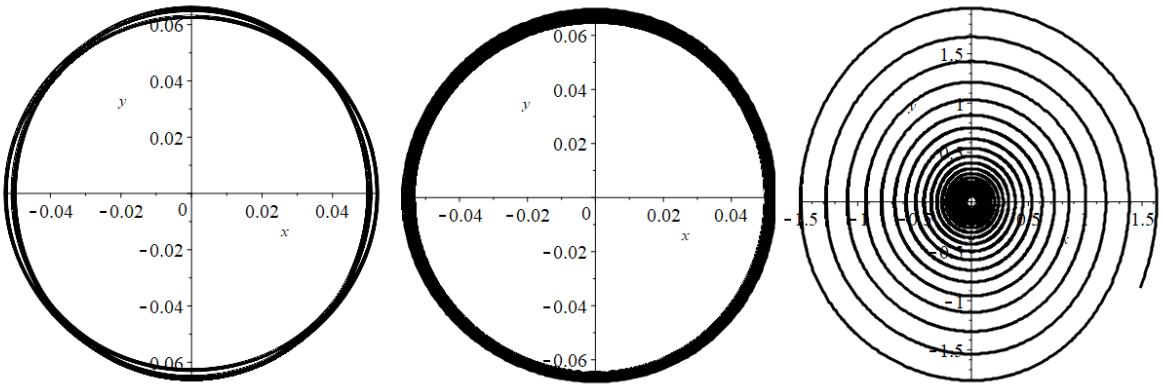
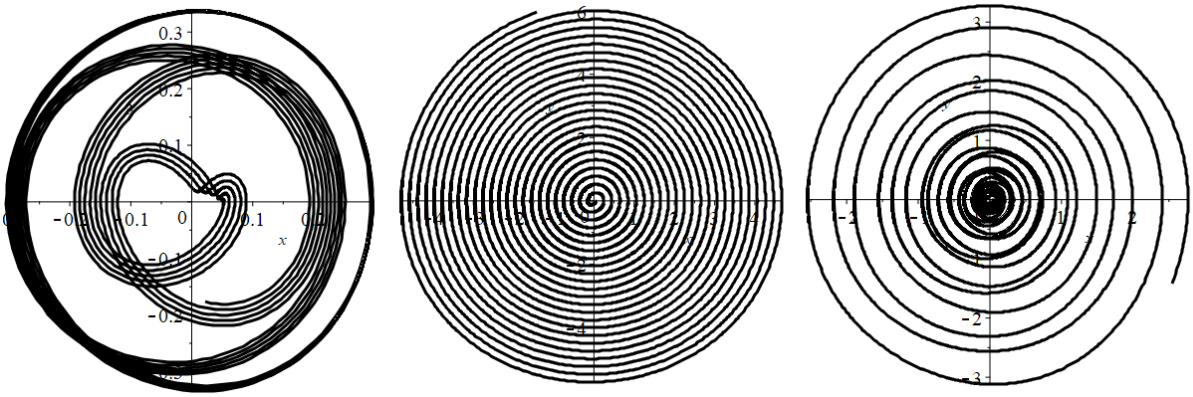


Рис.6.2. Фазові портрети руху рідини при вертикальних навантаженнях

Тут рис. А) відповідає $a_h = 0, a_v = 1,\ \omega_v = 1,$ а для рисунків В) та С) прийнято $a_h = 0,\ a_v = 1,\ \omega_v = 1.254$ та $a_h = 0, a_v = 1,\ \omega_v = 2.508$, відповідно. З наведених результатів бачимо, що в перших двох випадках рухи є стабільними, але при $\omega_v = 2.508$Гц відбувається необмежене зростання амплітуди, що відповідає випадку параметричного резонансу (частота сили, шо змушує, дорівнює подвійній фундаментальній частоті).

Далі розглянуті комбіновані вертикальне й горизонтальне навантаження, тобто додано горизонтальне навантаження. В результаті розрахунків отримані фазові портрети в координатах $(\zeta, \dot{\zeta})$, що наведені на рис.6.3.



. A B C

Рис.6.3. Фазові портрети руху рідини при комбінованих навантаженнях

Обрані такі параметри комбінованого навантаження: А) $a_h = 0.1,\ a_v = 1,\ \omega_h = \omega_v = 1$; В) $a_h = 0.1, a_v = 1,\ \omega_h = \omega_v = 1.254$; С) $a_h = 0.1, a_v = 1,\ \omega_h = \omega_v = 2.508$. Зауважимо, що в цьому випадку спостерігаємо появу ще одного резонансу, пов'язаного з горизонтальним навантаженням.

Зони стійкого руху рідини при лише вертикальних навантаженнях.3 можна з'ясувати за допомогою діаграми Айнса-Стретта [9], яка поділяє площину зміни параметрів навантаження на частини, що відповідають стабільним та нестабільним рухам, рис.6.4. Вводяться такі змінні

$$\kappa = \frac{\omega_{jk}^2}{\omega_v^2}, \quad \mu = \frac{a_v}{g}, \quad k = \overline{1,n},\ j = 0,1$$

для кожної фундаментальної частоти. Криві 1-5 на рис. 6.4 поділяють площину в координатах (κ, μ) на області, що відповідають нестійким рухам (темні області) та стійким рухам, які залишаються обмеженими в часі (білі області), зображено також точки А, В, С.

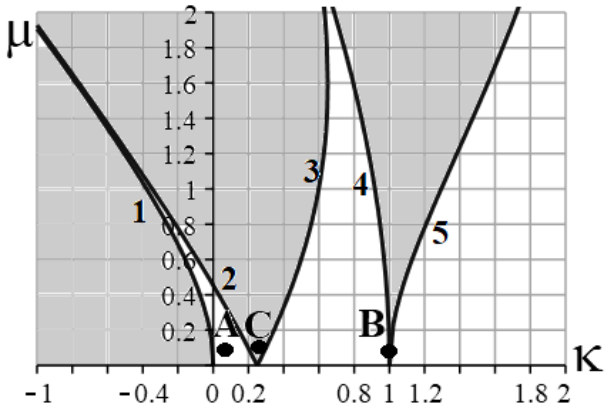


Рис.6.4. Діаграма Айнса - Стретта

Бачимо що точка А знаходиться в зоні стійкості точка В – на межі зони стійкості а точка С потрапляє в зону нестійкості руху.

Далі проаналізуємо рух рідини при дії комбінованого навантаження за умови наявності демпфування. Розв'яжемо систему диференціальних рівнянь 5.4 при таких початкових даних

$$d_{kl}(t) = 0,\ \dot{d}_{kl}(t) = 0,\ k = \overline{2,n},\ l \neq 1,\ \dot{d}_{11}(t) = 0.05.$$

Візьмемо коефіцієнт демпфування c=0.075. На рис. 6.5-6.7 наведені графіки зміни рівня вільної поверхні з часом за різні умови навантаження.

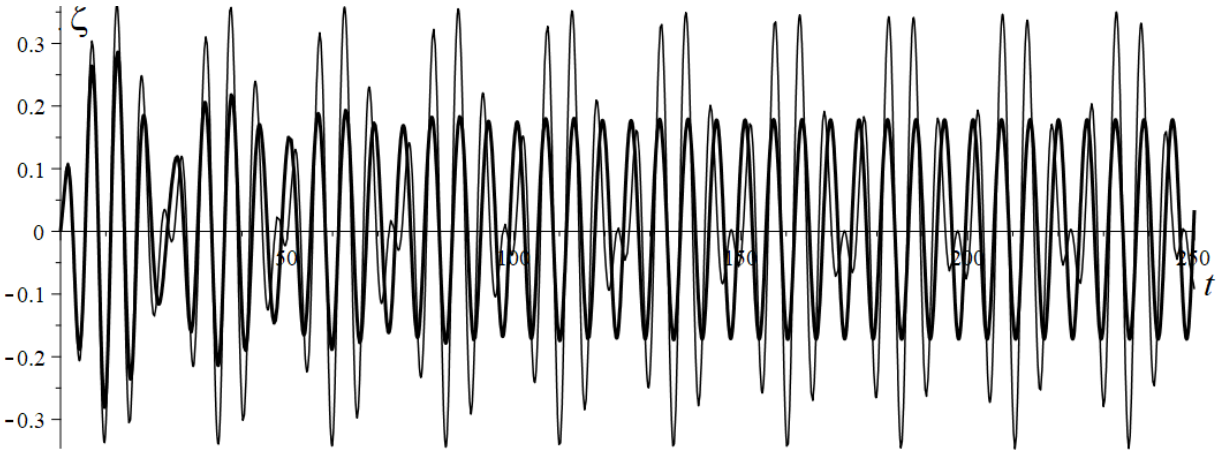


Рис.6.5. Зміна рівня вільної поверхні за часом у точці A

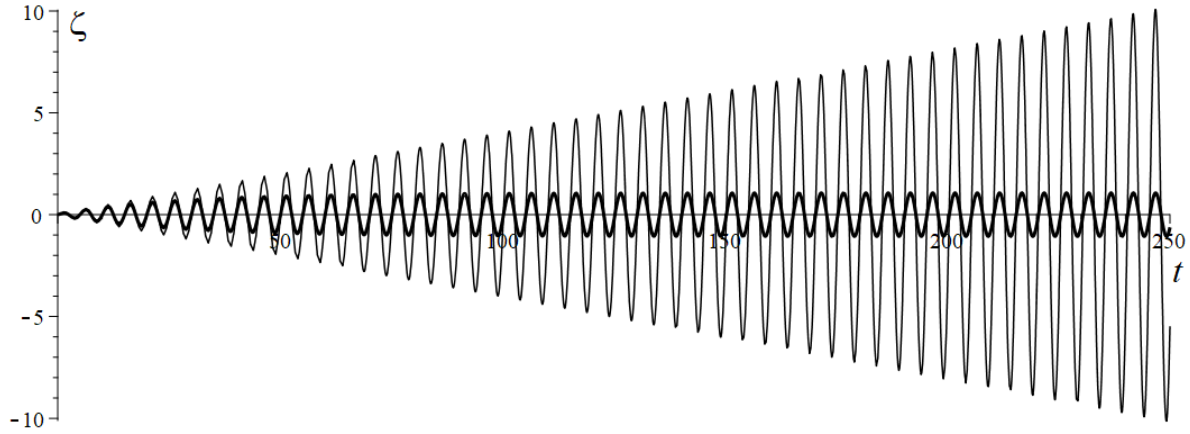


Рис.6.6. Зміна рівня вільної поверхні за часом у точці C

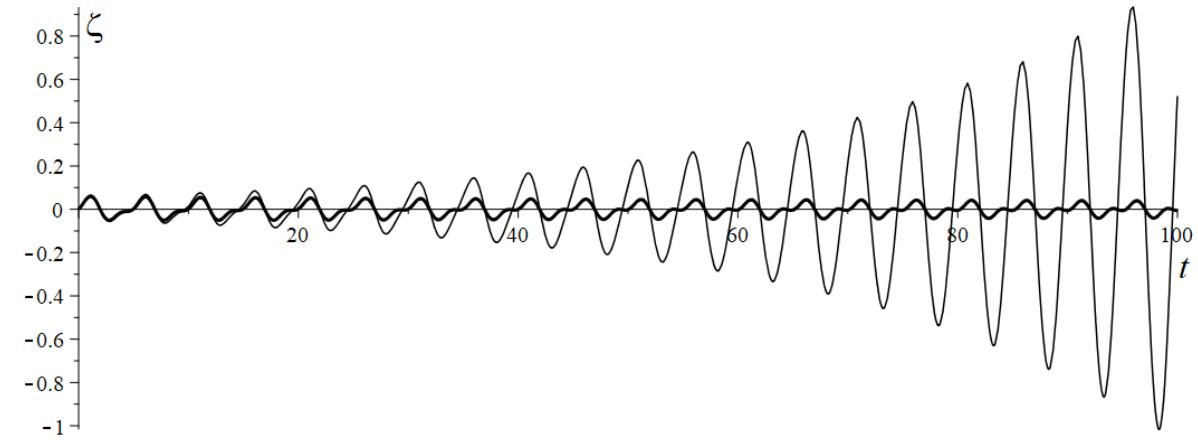


Рис.6.7. Зміна рівня вільної поверхні за часом у точці B

На рис. 6.5-6.7 зображено зміну рівня вільної поверхні у точці $(\theta = 0, r = R)$. Сірі лінії відповідають розв'язкам системи диференціальних рівнянь (5.2), тобто без врахування демпфування, чорні лінії отримані при розв'язанні системи диференціальних рівнянь (5.4) з урахуванням демпфування за Релеєм. Обраний коефіцієнт демпфування характеризує низький рівень демпфування, але у всіх розглянутих випадках спостерігається суттєве зменшення амплітуди коливань.

Висновки

Розроблено метод визначення змінного за часом рівня вільної поверхні рідини в жорстких оболонках обертання. Спектральна задача з визначення частот та форм коливань рідини в усіченому конічному резервуарі розв'язана шляхом зведення до системи одновимірних інтегральних рівнянь. За допомогою діаграми Айнса-Стретта знайдені зони нестійкості руху рідини при гармонічних вертикальних навантаженнях. З'ясовано вплив демпфування за Релеєм на зростання рівня вільної поверхні. В подальшому передбачається дослідження коливань пружних оболонок обертання з рідиною, з використанням різних композитних матеріалів [21].

СПИСОК ЛІТЕРАТУРИ

1. Karaiev A., Strelnikova E. Liquid Sloshing in Circular Toroidal and Coaxial Cylindrical Shells. *In: Ivanov, V., Pavlenko, I., Liaposhchenko, O., Machado, J., Edl, M. (eds) Advances in Design, Simulation and Manufacturing III. DSMIE 2020. Lecture Notes in Mechanical Engineering. Springer, Cham.* 2020. https://doi.org/10.1007/978-3-030-50491-5_1
2. Balas O.-M., Doicin C. V. and Cipu E. C. Analytical and Numerical Model of Sloshing in a Rectangular Tank Subjected to a Braking, *Mathematics*, vol. 11, pp. 949-955, 2023. DOI:10.3390/math11040949
3. Liu J., Zang Q., Ye W., Lin G. High performance of sloshing problem in cylindrical tank with various barrels by isogeometric boundary element method, *Engineering Analysis with Boundary Elements*, vol.114, pp.148-165, 2020. DOI:10.1016/j.enganabound.2020.02.014.
4. Krutchenko D. V., Strelnikova E. A., Shuvalova Y. S. Discrete singularities method in problems of seismic and impulse impacts on reservoirs. *Вісник Харківського національного університету імені В.Н.Каразіна, сер. «Математичне моделювання. Інформаційні технології. Автоматизовані системи управління»,* т. 35, С. 31-37, 2017, http://lib.kart.edu.ua/bitstream/123456789/13113/1/Krutchenko.pdf.
5. Lampart P., Rusanov A., Yershov S., Marcinkowski S., Gardzilewicz A. Validation of a 3D BANS solver with a state equation of thermally perfect and calorically imperfect gas on a multi-stage low-pressure steam turbine flow, *Journal of Fluids Engineering, Transactions of the ASME*, vol. 127(1), pp. 83–93,2005. DOI: 10.1115/1.185249.
6. Malykhina A., Merkulov D., Postnyi O., Smetankina N. Stationary problem of heat conductivity for complex-shape multilayer plates, *Вісник Харківського національного університету імені В.Н.Каразіна, сер. «Математичне моделювання. Інформаційні технології. Автоматизовані системи управління»,* т. 41, С. 46-54, 2019. DOI:10.26565/2304-6201-2019-41-05.
7. Murawski K. Technical Stability of Very Slender Rectangular Columns Compressed by Ball-And-Socket Joints without Friction, *Int. Journal of Structural Glass and Advanced Materials Research*, vol. 4(1), pp. 186-208, 2020. DOI: 10.3844/sgamrsp.2020.186.208
8. Tong C., Shao Y., Bingham H.B. & Hanssen, FC. W., An Adaptive Harmonic Polynomial Cell Method with Immersed Boundaries: Accuracy, Stability and Applications. *International Journal for Numerical Methods in Engineering*, vol. 122, pp. 2945–2980, 2021. https://doi.org/10.1002/nme.6648
9. Strelnikova E., Kriutchenko D., Gnitko V. Tonkonozhenko A., Liquid Vibrations in Cylindrical Tanks with and Without Baffles Under Lateral and Longitudinal Excitations. *International Journal of Applied Mechanics and Engineering*, vol. 25(3), pp.117-132, 2020. DOI:10.2478/ijame-2020-0038
10. Poguluri S. K., Cho Il H., Effect of vertical porous baffle on sloshing mitigation of two-layered liquid in a swaying tank, *Ocean Engineering*, vol. 289, Part 1, 2023,115952, https://www.sciencedirect.com/science/article/pii/S0029801823023363

11. Choudhary N., Bora S.N. and Strelnikova E., Study on liquid sloshing in an annular rigid circular cylindrical tank with damping device placed in liquid domain, *J. Vib. Eng. Tech.*, vol. 9, pp. 1–18, 2021. DOI:10.1007/s42417-021-00314-w
12. Choudhary N., Kumar N., Strelnikova E., Gnitko V., Kriutchenko D., Degtyariov K. Liquid vibrations in cylindrical tanks with flexible membranes. *Journal of King Saud University – Science*, vol. 33(8), 101589, 2021. doi.org/10.1016/j.jksus.2021.101589.
13. Sierikova O, Strelnikova E, Gnitko V, Degtyarev K., Boundary Calculation Models for Elastic Properties Clarification of Three-dimensional Nanocomposites Based on the Combination of Finite and Boundary Element Methods. IEEE 2nd KhPI Week on Advanced Technology (KhPIWeek), pp. 351–356, 2021. doi: 10.1109/KhPIWeek53812.2021.9570086
14. Konopka M., De Rose F., Strauch H., Jetzschmann C., Darkow N., Gerstmann J., Active slosh control and damping - Simulation and experiment, *Acta Astronautica*, vol. 158, pp. 89-102, 2019, https://doi.org/10.1016/j.actaastro.2018.06.055.
15. Zhang Y., Wan D., MPS-FEM coupled method for sloshing flows in an elastic tank, *Ocean Engineering*, vol. 152, pp. 416-427, 2018, https://doi.org/10.1016/j.oceaneng.2017.12.008.
16. Gnitko V., Karaiev A., Degtyariov K., Strelnikova E. Singular boundary method in a free vibration analysis of compound liquid-filled shells, *WIT Transactions on Engineering Sciences*, vol.126, pp.189-200, 2019. WIT Press, DOI:10.2495/BE420171.
17. Raynovskyy I. A. and Timokha A. N. Sloshing in Upright Circular Containers: Theory, Analytical Solutions, and Applications, 2020, CRC Press/Taylor and Francis Group, https://doi.org/10.1201/9780429356711
18. 1brahim R. A., 2005. Liquid Sloshing Dynamics. Theory and Applications. Cambridge University Press.
19. Pradeepkumar K., Selvan V., Satheeshkumar K., Review of Numerical Methods for Sloshing. *International Journal for Research in Applied Science & Engineering Technology*. vol.8, Issue XI, 2020, doi.org/10.22214/ijraset.2020.32116.
20. Gavrilyuk I., Hermann M., Lukovsky I., Solodun O., Timokha A., Natural Sloshing frequencies in Truncated Conical Tanks, *Engineering Computations*, vol. 25, no. 6, pp.518 – 540, 2008, DOI: 10.1108/02644400810891535.
21. Sierikova O., Strelnikova E. and Degtyariov K., Srength Characteristics of Liquid Storage Tanks with Nanocomposites as Reservoir Materials, *2022 IEEE 3rd KhPI Week on Advanced Technology (KhPIWeek), Kharkiv, Ukraine*, pp. 1-7, 2022, doi: 10.1109/KhPIWeek57572.2022.9916369

REFERENCES

1. A. Karaiev, E. Strelnikova, (2020). Liquid Sloshing in Circular Toroidal and Coaxial Cylindrical Shells. In: Ivanov, V., Pavlenko, I., Liaposhchenko, O., Machado, J., Edl, M. (eds) Advances in Design, Simulation and Manufacturing III. DSMIE 2020. Lecture Notes in Mechanical Engineering. Springer, Cham. https://doi.org/10.1007/978-3-030-50491-5_1
2. O.-M. Balas C. V. Doicin and E. C. Cipu, (2023). Analytical and Numerical Model of Sloshing in a Rectangular Tank Subjected to a Braking, *Mathematics*, Vol. 11, P. 949-955, DOI:10.3390/math11040949
3. J. Liu Q. Zang, W. Ye, G. Lin, (2020). High performance of sloshing problem in cylindrical tank with various barrels by isogeometric boundary element method", *Engineering Analysis with Boundary Elements*, , Vol. 114, P. 148-165, DOI:10.1016/j.enganabound.2020.02.014
4. D. V. Krutchenko, E. A. Strelnikova, Shuvalova Y. S. (2017). Discrete singularities method in problems of seismic and impulse impacts on reservoirs. *Bulletin of VN Karazin Kharkiv National University, series «Mathematical modeling. Information technology. Automated control systems»*, vol. 35, pp. 31-37. http://lib.kart.edu.ua/bitstream/123456789/13113/1/Krutchenko.pdf.
5. P. Lampart, A. Rusanov, S. Yershov, S. Marcinkowski, A. Gardzilewicz, (2005).Validation of a 3D BANS solver with a state equation of thermally perfect and calorically imperfect gas on a multi-stage low-pressure steam turbine flow, *Journal of Fluids Engineering, Transactions of the ASME*, vol. 127(1), pp. 83–93,2005. DOI: 10.1115/1.185249.
6. A. Malykhina, D. Merkulov, O. Postnyi, N. Smetankina, (2019). Stationary problem of heat conductivity for complex-shape multilayer plates. *Bulletin of VN Karazin Kharkiv National*

University, series «Mathematical modeling. Information technology. Automated control systems», vol. 41, pp. 46-54,. DOI:10.26565/2304-6201-2019-41-05.

7. K.Murawski, (2020). Technical Stability of Very Slender Rectangular Columns Compressed by Ball-And-Socket Joints without Friction, *Int. Journal of Structural Glass and Advanced Materials Research*, vol, 4(1), pp. 186-208, DOI: 10.3844/sgamrsp.2020.186.208
8. C.Tong, Y. Shao, H. B. Bingham, & FC. W. Hanssen, (2021). An Adaptive Harmonic Polynomial Cell Method with Immersed Boundaries: Accuracy, Stability and Applications. *International Journal for Numerical Methods in Engineering*, , Vol. 122, P. 2945–2980. https://doi.org/10.1002/nme.6648.
9. E. Strelnikova, D. Kriutchenko, V. Gnitko, A. Tonkonozhenko, (2020).Liquid Vibrations in Cylindrical Tanks with and Without Baffles Under Lateral and Longitudinal Excitations, *International Journal of Applied Mechanics and Engineering*, Vol. 25, Issue 3, P. 117-132, DOI: 10.2478/ijame-2020-0038.
10. S. K. Poguluri, Il H. Cho, (2023).Effect of vertical porous baffle on sloshing mitigation of two-layered liquid in a swaying tank, *Ocean Engineering*, vol. 289, Part 1, 115952, https://www.sciencedirect.com/science/article/pii/S0029801823023363
11. N. Choudhary, S.N. Bora and E. Strelnikova, (2021). Study on liquid sloshing in an annular rigid circular cylindrical tank with damping device placed in liquid domain, *J. Vib. Eng. Tech.*, vol. 9, pp. 1–18, DOI:10.1007/s42417-021-00314-w
12. N. Choudhary, N. Kumar, E. Strelnikova, V. Gnitko, D. Kriutchenko, K. Degtyariov, (2021). Liquid vibrations in cylindrical tanks with flexible membranes. *Journal of King Saud University – Science*, vol. 33(8), 101589, doi.org/10.1016/j.jksus.2021.101589.
13. O. Sierikova, E. Strelnikova, Gnitko V, Degtyarev K. (2021).Boundary Calculation Models for Elastic Properties Clarification of Three-dimensional Nanocomposites Based on the Combination of Finite and Boundary Element Methods. *IEEE 2nd KhPI Week on Advanced Technology (KhPIWeek)*, pp. 351–356,.doi: 10.1109/KhPIWeek53812.2021.9570086
14. M. Konopka, F., De Rose, H. Strauch, C. Jetzschmann, N. Darkow, J. Gerstmann, (2019). "Active slosh control and damping - Simulation and experiment, *Acta Astronautica*, vol. 158, pp. 89 - 102, https://doi.org/10.1016/j.actaastro.2018.06.055.
15. Y. Zhang, D. Wan, (2018), MPS-FEM coupled method for sloshing flows in an elastic tank", *Ocean Engineering*, vol. 152, pp. 416-427, https://doi.org/10.1016/j.oceaneng.2017.12.008
16. V. Gnitko, A. Karaiev, K.Degtyariov, E.Strelnikova, (2019). Singular boundary method in a free vibration analysis of compound liquid-filled shells, *WIT Transactions on Engineering Sciences*, Vol. 126, P. 189-200, WIT Press, DOI:10.2495/BE420171.
17. I. A. Raynovskyy and A. N. Timokha, (2020). Sloshing in Upright Circular Containers: Theory, Analytical Solutions, and Applications, CRC Press/Taylor and Francis Group, https://doi.org/10.1201/9780429356711
18. R. A. 1brahim, *Liquid Sloshing Dynamics. Theory and Applications*. Cambridge University Press. 2005, 984 p.
19. K. Pradeepkumar, V. Selvan, K.Satheeshkumar, (2020). Review of Numerical Methods for Sloshing, *International Journal for Research in Applied Science & Engineering Technology*, Vol. 8, Issue XI, doi.org/10.22214/ijraset.2020.32116.
20. Gavrilyuk I., Hermann M., Lukovsky I., Solodun O., Timokha A. (2008). Natural Sloshing frequencies in Truncated Conical Tanks, Engineering Computations, vol. 25, no. 6, pp. 518 – 540, DOI: 10.1108/02644400810891535
21. O. Sierikova, E. Strelnikova and K. Degtyariov, (2022). Srength Characteristics of Liquid Storage Tanks with Nanocomposites as Reservoir Materials, *2022 IEEE 3rd KhPI Week on Advanced Technology (KhPIWeek), Kharkiv, Ukraine*, pp. 1-7, DOI:10.1109/KhPIWeek57572.2022.9916369.

Degtyarev Kirill *PhD, researcher*
A. Pidhorny Institute of Mechanical Engineering Problems
vul. Pozharskogo, 2/10, Kharkiv, 61046, Ukraine
Kolodiazhny Andriy *Post-graduate student*
A. Pidhorny Institute of Mechanical Engineering Problems
vul. Pozharskogo, 2/10, Kharkiv, 61046, Ukraine
Kriutchenko Denys *PhD, junior researcher*
A. Pidhorny Institute of Mechanical Engineering Problems
vul. Pozharskogo, 2/10, Kharkiv, 61046, Ukraine
Usaova Olga *engineer*
A. Pidhorny Institute of Mechanical Engineering Problems
vul. Pozharskogo, 2/10, Kharkiv, 61046, Ukraine
Strelnikova Olena *DSc, Prof., Leading researcher*
A.M. Pidhorny Institute for Mechanical Engineering Problems of the Ukrainian Academy of Sciences, 2/10, Communalnikiv Str., Kharkiv, 61046

Computer modelling of liquid sloshing in tanks under periodic loads

The main purpose of the work is to develop the computer methodology for taking damping into account when analysing the stability of fluid movement in reservoirs and fuel tanks under periodic external loads

Relevance. Damping plays a critical role in providing stability and reducing potential hazards in tanks which are partially filled with liquid. Lack of damping can lead to motion instability. In reservoirs filled with liquid, any movement disturbances such as sudden acceleration, deceleration or turning can cause sloshing. Without damping, sloshing can grow, potentially leading to uncontrolled and dangerous situations, especially in vehicles or at industrial processes. Damping provides control over the sloshing dynamics, providing smoother and more predictable behaviour. By damping excessive vibrations, engineers can ensure that the fluid remains stable inside the fuel tank, reducing the risk of excessive dynamic loads on the tank structure or the vehicle carrying it. Therefore, studies devoted to the study of damping are highly relevant.

Research methods. The methods of integral equations, the method of given forms, and the method of boundary elements have been used to solve the problem of damping.

The results. The spectral boundary value problem has been solved and the frequencies and forms of natural oscillations of the fluid have been found. Combined horizontal and vertical loads have been studied, and zones of stable and unstable movement which are depending on the load parameters have been found. The effect of damping has been studied by using the Rayleigh matrix. The importance of the obtained results on fluid splashing in rigid tanks lies in clarifying the critical role of damping in providing stability and reducing potential hazards to the stability of launch vehicle fuel tanks during flight.

Conclusions. The method for determining the time-varying level of the liquid free surface in rigid shells of revolution has been developed. The spectral problem of determining the frequencies and modes of liquid oscillations in a truncated conical tank is solved by reducing it to the system of one-dimensional integral equations. With the help of the Ince-Strutt diagram, the zones of instability of fluid movement under harmonic vertical loads have been found. The effect of Rayleigh damping on the growth of the free surface level has been clarified. In the future, it is planned to study the oscillations of elastic shells of rotation with liquid, using various composite materials

Keywords: *free surface, liquid sloshing in tanks, systems of singular integral equations, boundary element method, damping, Ince-Strutt diagram.*

Надійшла у першій редакції 11.03.2024, в останній - 17.04.2024.
The first version has been received on 11.03.2024, the final version - on 17.04.2024.

DOI: https://doi.org/10.26565/2304-6201-2024-61-03

УДК 681.5.015

Дядюн Сергій Васильович
к.т.н., доцент;доцент кафедри моделювання систем і технологій;
Харківський національний університет імені В.Н. Каразіна
Майдан Свободи 4, Харків, Україна 61022
e-mail: s.v.dyadun@karazin.ua
https://orcid.org/0000-0002-1910-8594

Аналіз ефективності методів оптимізації потокорозподілу в системах водопостачання з великою кількістю насосних станцій

Актуальність. В даний час досліджено методи оптимізації для невеликої кількості працюючих на мережу активних джерел. Однак при розробці систем оперативного управління режимами функціонування систем подачі та розподілу води (СПРВ) великих міст доводиться стикатися з великою кількістю насосних станцій (НС), що одночасно працюють на мережу. Складність розв'язання задачі оптимізації потокорозподілу в СПРВ зростає зі збільшенням кількості спільно працюючих активних джерел, які є змінними оптимізації задачі, що розглядається.

Мета. В проблемі оперативного управління режимами функціонування СПРВ важливе місце посідає задача оптимізації потокорозподілу у водопровідній мережі великої розмірності. Мета задачі - необхідно так розподілити навантаження (витрати) між НС, щоб при забезпеченні заданої якості постачання водою усіх споживачів досягався мінімум суми енерговитрат на насосних станціях. Розглянуто постановку задачі, методи її вирішення для СПРВ великого міста, на яку працює велика кількість насосних станцій. Необхідно провести порівняльний аналіз ефективності використання різних оптимізаційних методів для вирішення задачі оптимального розподілу навантаження між великою кількістю НС, одночасно працюючих на систему водопостачання мегаполісу.

Методи дослідження. Дану задачу можна вирішити методами нелінійного математичного програмування чи пошукової оптимізації на базі гідравлічного розрахунку водопровідної мережі. Її специфічна особливість – алгоритмічне завдання функції мети. При роботі на мережу двох активних джерел така задача зводиться до задачі одновимірної пошукової оптимізації. При більшій кількості змінних необхідно використовувати методи багатовимірної оптимізації. Для дослідження ефективності вирішення задачі оптимізації потокорозподілу в СПРВ використовувалися найбільш ефективні та поширені методи: покоординатного спуску; сканування зі змінним кроком; деформованого багатогранника Нелдера-Міда; прямого пошуку Хука та Дживса; Розенброка; Пауелла.

Результати. Проведені дослідження показали, що найбільш ефективним за критеріями мінімуму витрат комп'ютерного часу та об'єму комп'ютерної пам'яті виявився метод прямого пошуку Хука і Дживса.

Висновки. Отримані результати доцільно використовувати для розробки та експлуатації систем оперативного управління режимами функціонування СПРВ великих міст, АРМ диспетчерів водопровідних мереж, САПР систем водопостачання для визначення оптимальних режимів функціонування СПРВ.

Ключові слова: математична модель, система водопостачання, насосна станція, функціонування, потокорозподіл, оперативне управління, критерій, метод, ефективність.

Як цитувати: Дядюн С. В. Аналіз ефективності методів оптимізації потокорозподілу в системах водопостачання з великою кількістю насосних станцій. *Вісник Харківського національного університету серія Математичне моделювання. Інформаційні технології. Автоматизовані системи управління*. 2024. вип. 61. С. 25–32.
https://doi.org/10.26565/2304-6201-2024-61-03

How to quote: Dyadun S. V., "Analysis of the effectiveness of flow distribution optimization methods in water supply systems with a large number of pumping stations", *Bulletin of V. Karazin Kharkiv National University series Mathematical Modeling. Information Technology. Automatized Control Systems,* 2024, Issue 61, pp. 25–32. [In Ukrainian] https://doi.org/10.26565/2304-6201-2024-61-03

1 Вступ

В даний час в достатній мірі досліджено і апробовано на практиці методи оптимізації для невеликої кількості працюючих на мережу активних джерел [1, 2, 22]. Однак при розробці систем оперативного управління технологічними процесами функціонування систем подачі та розподілу

води (СПРВ) великих міст і автоматизованих робочих місць (АРМ) диспетчерів водопровідних мереж доводиться стискатися з великою кількістю насосних станцій, що одночасно працюють на мережу. Складність розв'язання задачі оптимізації потокорозподілу в СПРВ зростає зі збільшенням кількості спільно працюючих активних джерел, які є змінними оптимізації задачі, що розглядається.

Математичному моделюванню систем водопостачання та інших трубопровідних систем присвячено велику кількість робіт, серед яких хотілося б виділити [1-3, 7, 8, 9, 10, 12, 14, 20], оперативному плануванню роботи СПРВ [1-3, 11, 14, 16], оптимізації режимів роботи насосних станцій СПРВ [13, 15, 17, 18].

У даній статті проводиться порівняльний аналіз ефективності використання різних оптимізаційних методів для визначення оптимального потокорозподілу в системах водопостачання з великою кількістю насосних станцій.

2 Постановка проблеми

Системи подачі та розподілу води сучасних міст відносяться до класу великих систем, управління якими можливе лише в умовах неповної та недостовірної інформації про керований об'єкт.

У роботах [1-3] при оперативному управлінні СПРВ запропоновано і знайшов широке застосування на практиці підхід, суть якого полягає у використанні специфічних властивостей водопровідних мереж, пов'язаних з наявністю в них диктуючих точок, і полягає в тому, що якість функціонування всієї мережі можна характеризувати її станом у цих точках. Диктуючою точкою водопровідної мережі в даний час є така точка, в якій величина тиску в той же момент часу мінімальна. Стохастичний характер процесів водоспоживання викликає безперервну зміну потокорозподілу в мережі та появу деякої множини диктуючих точок. Використання цих властивостей дозволяє синтезувати систему управління, яка б забезпечувала задану якість функціонування СПРВ в диктуючих точках.

Вирішення проблеми оперативного управління потокорозподілом у системах подачі та розподілу води на кожному з етапів - оперативного планування та стабілізації - як правило, рознесене в часі та просторі, вимагає різного обсягу, складу та характеру оперативної інформації, різних математичних моделей, що описують об'єкт управління, критеріїв та методів вирішення задач управління. У той же час здійснення кожного з етапів управління забезпечується в результаті вирішення певного ряду задач. Ці задачі можна розділити на задачі моделювання, які вирішуються поза контуром реального управління, і безпосередньо задачі управління, які вирішуються в оперативному режимі.

Для вирішення задач етапу оперативного планування процесів потокорозподілу в СПРВ найбільш адекватною математичною моделлю об'єкта управління є модель потокорозподілу, що встановився. В основі моделі такого потокорозподілу містяться відомості про топологію мережі та технологічні параметри всіх елементів еквівалентної схеми водопровідної мережі. Для вирішення задач стабілізації режимів використовуються динамічні моделі [1-3].

Відповідно до такого підходу процес оперативного управління потокорозподілом у СПРВ розбивається на два етапи: оперативне планування режимів функціонування СПРВ та їх стабілізація [1-3]. Задача оперативного управління СПРВ формулюється як двоетапна задача нелінійного стохастичного програмування [2].

На верхньому рівні етапу оперативного планування режимів функціонування СПРВ вирішується задача оптимізації режимів функціонування насосних станцій при їх спільній роботі на мережу. У якості критерія використовується мінімум суми енерговитрат на насосних станціях. Як результат її вирішення визначаються значення витрат і тисків на виходах насосних станцій, які потрібно запланувати, а також місце розташування диктуючих на майбутній момент часу точок водопровідної мережі, в яких необхідно стабілізувати такі значення тиску [1,2,14]. Далі на нижньому рівні етапу оперативного планування режимів з метою забезпечення запланованих параметрів на виходах насосних станцій для кожної з них визначаються оптимальні структури та параметри її функціонування [13]. Стабілізація тисків у диктуючих точках мережі дозволяє скоригувати помилки планування шляхом уточнення тисків у цих точках.

3 Математична постановка задачі

Таким чином, у проблемі оперативного управління технологічними процесами функціонування систем подачі та розподілу води важливе місце посідає задача оптимізації потокорозподілу у водопровідній мережі. Розглянемо постановку задачі, методи її вирішення для СПРВ великого міста, на яку працює велика кількість насосних станцій (НС).

Мета цієї задачі полягає у визначенні таких значень компонент векторів витрати $\overline{q}^{(a)}_{вых}$ і тиску $\overline{H}^{(a)}_{вых}$ на виході всіх насосних станцій, які при відомих прогнозних значеннях вузлових витрат q_j, j∈N в мережі та виконанні всіх обмежень, що накладаються, забезпечували б досягнення максимуму деякого критерію, який характеризує ефективність функціонування СПРВ, тобто необхідно так розподілити навантаження (витрати) між НС, щоб при забезпеченні заданої якості постачання водою всіх споживачів досягався максимум ефективності.

Будемо використовувати у якості критерію ефективності функціонування СПРВ мінімум суми енерговитрат на насосних станціях. Позначимо L, M, N - відповідно безлічі активних елементів, магістральних ділянок мережі та фіктивних дуг зі споживачами; E = LUMUN. Індекс 1, відповідний елементам цих множин, характеризує їхню приналежність до гілок дерева графа мережі, індекс 2 - до хорд. Тоді задача оптимізації розподілу навантаження між СПРВ при їх спільній роботі на водопровідну мережу буде мати вигляд

$$\Phi\left[\overline{q}^{(a)}_{вых}, \overline{H}^{(a)}_{вых}\right] = \sum_{i\in L} P_i^*(q^{(a)}_{выхi}, H^{(a)}_{выхi}) \to \min_{q^{(a)}_{выхi}\in\Omega} \tag{1}$$

$$\Omega: \quad f_r = signq_r r_r |q_r|^2 + \sum_{i\in M_1} b_{1ri} signq_i r_i |q_i|^2 = 0, r \in M_2 \tag{2}$$

$$f_r = H_{rh} - H_{jk} + \sum_{i\in M_1} b_{1ri}(signq_i r_i |q_i|^2 + h_i^{(\Gamma)}) = 0, r \in N,\ j \in L_1 \tag{3}$$

$$f_r = H_{jk} - H_{rh} + \sum_{i\in M_1} b_{1ri}(signq_i r_i |q_i|^2 + h_i^{(\Gamma)}) = 0, r \in L_2,\ j \in L_1 \tag{4}$$

$$q_i = \sum_{r\in M_2 UL_2} b_{1ri} q_r + Q_i^+, i \in M_1 UN_1 \tag{5}$$

$$q^{(a)+}_{выхi}(H^{(a)}_{выхi}) \le q^{(a)}_{выхi} \le q^{(a)++}_{выхi}(H^{(a)}_{выхi}), i \in L \tag{6}$$

$$h_j \ge h_j^+, j \in N, \tag{7}$$

де

$$P_i^*(\cdot) = q^{(a)*}_{выхi} H^{(a)*}_{выхi}, i \in L; \tag{8}$$

$$Q_i^+ = \sum_{r\in N} b_{1ri} q_r^+ = const;$$

$q^{(a)*}_{выхi}, H^{(a)*}_{выхi}$ - оптимальні значення витрати та тиску на виході і-ї НС; b_{1ri} - елемент цикломатичної матриці $B_1[I]$; $q_i, i\epsilon E$ - витрата води в і-й дузі водопровідної мережі; $r_i, i\epsilon M$ - гідравлічний опір і-ї ділянки СПРВ; $h_i^{(\Gamma)}, i\epsilon M$ - перепад геодезичних висот і-ї ділянки СПРВ; $q^{(a)+}_{выхi}, q^{(a)++}_{выхi}, i\epsilon L$ - нижня і верхня межі значення витрати через НС, що залежать від величини $H^{(a)}_{выхi}; h_j, h_j^+, j\epsilon N$ - тиск у j-му вузлі СПРВ, відповідно фактичний та мінімально допустимий, H_{rh}, H_{jk} - тиски відповідно на початку r-ї та в кінці j-ї дуги.

4 Методи вирішення

Дану задачу можна вирішити методами нелінійного математичного програмування чи пошукової оптимізації на базі гідравлічного розрахунку водопровідної мережі. Її специфічна особливість – алгоритмічне завдання функції мети.

Замість розв'язання складної задачі математичного програмування (1)-(8) доцільно в області

$$q_{выхi}^{(a)+}(H_{выхi}^{(a)}) \le q_{выхi}^{(a)} \le q_{выхi}^{(a)++}(H_{выхi}^{(a)}), i \in L; \quad (9)$$

$$\sum_{i \in L} q_{выхi}^{(a)} = \sum_{j \in N} q_j^{(н)} \quad (10)$$

шукати таку точку $\bar{q}_{выхi}^{(a)} = \left[q_{выхi}^{(a)}, i \in L\right]$, у якій після розв'язання задачі гідравлічного розрахунку значення критерію (1) найменше.

Як показали проведені дослідження, функція $\Phi\left[\bar{q}_{вых}^{(a)}, \bar{H}_{вых}^{(a)}\right]$, яка визначається відповідно до виразу (1), завжди унімодальна і опукла вниз. Цікавою властивістю цієї задачі є те, що функція $\Phi[\cdot]$ поза допустимою областю не визначена. У цьому випадку достатньо замінити значення цільової функції в неприпустимих точках дуже великою позитивною величиною, і вони автоматично відкидатимуться в процесі пошуку [5, 6].

Найпростішим методом розв'язання задачі оптимізації потокорозподілу в СПРВ є випадковий пошук в області Ω. Для мережі, на яку працює один активний елемент, розв'язання задачі гідравлічного розрахунку буде й оптимальним за критерієм (1). При роботі на мережу двох активних джерел така задача зводиться до задачі одновимірної пошукової оптимізації. За більшої кількості змінних необхідно використовувати методи багатовимірної оптимізації.

Для дослідження ефективності вирішення задачі оптимізації потокорозподілу в СПРВ будемо використовувати також найбільш ефективні та поширені методи: покоординатного спуску; сканування зі змінним кроком; деформованого багатогранника Нелдера-Міда; прямого пошуку Хука та Дживса; Розенброка; Пауелла [4,6,21]. Оскільки функція мети задана алгоритмічно, не вдається отримати аналітичні вирази для її похідних. Тому застосування методів багатовимірної оптимізації більш високих порядків не уявляється можливим.

5 Результати досліджень

В якості критеріїв ефективності при порівнянні цих методів використовувалися витрати машинного часу, необхідного для досягнення збіжності алгоритмів, та обсяг пам'яті ПК, що займається. Дослідження проводилися на основі моделі реальної СПРВ, яка складається з 382 дуг мережі та 10 активних джерел. При імітаційному моделюванні кількість спільно працюючих на водопровідну мережу активних джерел варіювалася в межах $3 \le l \le 10$, відповідно кількість незалежних змінних $l-1$.

За обсягом пам'яті ПК відмінності виявилися несуттєвими. В основному вони визначаються числом змінних l функції, що мінімізується, тоді як для обчислення функції мети використовуються масиви значно більшої розмірності.

Таким чином, будемо вважати основним критерій мінімуму витрат машинного часу.

Метод сканування зі змінним кроком найбільш простий у реалізації і дозволяє при досить густому розташуванні досліджуваних точок завжди знаходити глобальний екстремум. Однак, незважаючи на його надійність, він виявляється неефективним через надзвичайно велику кількість обчислень функції, що різко зростає зі збільшенням кількості змінних.

Недоліком методу покоординатного спуску є суттєва залежність його швидкості збіжності від вибору системи координат. Метод деформованого багатогранника недостатньо надійний. Метод Хука та Дживса працює по гребенях, його недолік полягає в тому, що можлива перепустка гребеня. Існують й інші методи мінімізації, які не потребують обчислення похідних цільової функції. Але більшість їх використовує різницеві апроксимації приватних похідних функцій $f(x)$, тобто, сутнісно є варіантами градієнтних методів.

Результати чисельного аналізу при роботі на мережу трьох активних джерел свідчать про гарну збіжність всіх методів при відносно невеликій кількості ітерацій (k <30). Метод деформованого багатогранника дещо поступається іншим методам. Метод покоординатного спуску швидко збігається на першому етапі визначення напрямку мінімізації, потім при роботі методів одновимірної мінімізації за напрямом його ефективність знижується. В процесі аналізу в якості процедури мінімізації за напрямом використовувалося кілька методів (метод золотого перерізу, метод Фібоначчі, обчислення мінімуму за допомогою квадратичної апроксимації з локалізацією точки мінімуму, метод дихотомії). Усі вони дають практично однакову збіжність.

Ефективність алгоритмів оптимізації потокорозподілу в СПРВ залежить від розмірності задачі, тобто від кількості $l-1$ незалежних змінних. У деяких з методів, що розглядалися, на кожній ітерації виконується декілька обчислень функції. У табл.1 наведено кількість обчислень функції, необхідну для досягнення відносної похибки обчислення, яка дорівнює 0,005%, для різного числа змінних.

Таблиця 1.

Метод	Число змінних		
	2	5	9
Прямий пошук Хука і Дживса	23	136	230
Покоординатний спуск	22	498	1000
Деформований багатогранник Нелдера-Міда	29	1000	1000
Розенброка	17	174	348
Пауелла	18	209	303

Переваги методу Хука та Дживса очевидні. Навіть при 9 змінних він дає цілком прийнятну для практичних цілей швидкість збіжності, тоді як при роботі методу багатогранника, що деформується, необхідний ступінь точності не був досягнутий і при 1000 обчисленнях функції навіть для 5 змінних. Збіжність методу покоординатного спуску зі збільшенням числа змінних також погіршується. Це пов'язано з тим, що на кожній ітерації відбувається n звернень до підпрограми обчислення мінімуму функції за напрямом, яка вимагає до 10 обчислень функції (хоча кількість ітерацій під час роботи методів Хука і Дживса та покоординатного спуску відрізняється незначною мірою).

У табл.2 наведено кількість ітерацій, необхідних для досягнення заданої точності збіжності розглянутих алгоритмів, при різному числі змінних оптимізації.

Таблиця 2.

Метод	Число змінних	
	2	5
Прямий пошук Хука і Дживса	2-3	10-12
Покоординатний спуск	2-3	10-12
Деформований багатогранник Нелдера-Міда	15-17	100
Розенброка	2-3	10-15
Пауелла	2-3	10-15

6 Висновки

Таким чином, при вирішенні задач оптимізації потокорозподілу в системах водопостачання великої розмірності (кількість активних джерел, що працюють на мережу, $3 \le l \le 10$) найбільш ефективним за критеріями витрат машинного часу і обсягу пам'яті ПК, що займається, є метод прямого пошуку Хука і Дживса.

Отримані результати доцільно використовувати для розробки та експлуатації систем оперативного управління технологічними процесами функціонування СПРВ великих міст, АРМ диспетчерів водопровідних мереж, САПР систем водопостачання для визначення оптимальних режимів функціонування СПРВ.

СПИСОК ЛІТЕРАТУРИ

1. Евдокимов А.Г., Тевяшев А.Д. Оперативное управление потокораспределением в инженерных сетях. Харьков, 1980, 144с.
2. Евдокимов А.Г., Тевяшев А.Д., Дубровский В.В. Моделирование и оптимизация потокораспределения в инженерных сетях. М: Стройиздат, 1990, 368с.
3. Евдокимов А.Г., Дубровский В.В., Тевяшев А.Д. Потокораспределение в инженерных сетях. М.: Стройиздат, 1979, 199с.

4. Химмельблау Д. Прикладное нелинейное программирование. М., 1975, 534с.
5. Растригин Л.А. Системы экстремального управления. М., 1974, 630с.
6. Евдокимов А.Г. Минимизация функций и ее приложения к задачам автоматизированного управления инженерными сетями. Харьков, 1985, 288с.
7. Fallside, F., Perry, P.F., Burch, R.H., Marlow, K.C.: The Development of Modelling and Simulation Techniques Applied to a Computer - Based - Telecontrol Water Supply System. In: Computer Simulation of Water Resourses Systems, 1975, No.12, pp. 617-639.
8. Абрамов Н.Н. Теория и методика расчета систем подачи и распределения воды. М.: Стройиздат, 1985, 288с.
9. Меренков А.П., Хасилев В.Я. Теория гидравлических цепей. М.: Наука, 1985, 279 с.
10. Новицкий Н.Н., Сухарев М.Г., Тевяшев А.Д. и др. Трубопроводные системы энергетики. Методические и прикладные проблемы математического моделирования. Изд-во ”Наука”, Новосибирск, 2015, 476с. http://51.isem.irk.ru/semtps/works.php
11. Tevyashev, A.D., Matvienko, O.I. About one approach to solve the problem of management of the development and operation of centralized water–supply systems. In: Econtechmod. An International Quarterly Journal. 2014, Vol. 3, Issue 3, pp. 61–76. https://openarchive.nure.ua/server/api/core/bitstreams/d73ad31a-bb66-46ac-987f-14a1943d1b1b/content
12. Novitsky N.N., Vanteyeva O.V. Modeling of stochastic hydraulic conditions of pipeline systems // Chaotic Modeling and Simulation (CMSIM). 2014. No.1. P. 95-108. http://www.asmda.es/images/Corrected-BOOK_OF_ABSTRACTS-ASMDA2017-5-22.pdf
13. Дядюн С.В. Выбор оптимальных комбинаций агрегатов насосной станции городского водопровода. // Коммунальное хозяйство городов, Киев, Техніка, 1992, №1, с.63-70.
14. Дядюн С.В. Моделирование и рациональное управление системами водоснабжения при минимальном объёме оперативной информации // Радиоэлектроника и информатика, Харьков, ХНУРЭ, № 20, 2002, с.111-115. https://cyberleninka.ru/article/n/modelirovanie-i-ratsionalnoe-upravlenie-sistemami-vodosnabzheniya-pri-minimalnom-obeme-operativnoy-informatsii
15. Reinbold, C., Hart, V. The search for energy savings: optimization of existing & new pumping stations. In: Florida Water Resources Journal, 2011, pp. 44–52.
16. Burgschweiger, J., Gnadig, B.B., Steinbach, M.C. Nonlinear programming techniques for operative planning in large drinking water networks. In: Konrad–Zuse–Zentrum fur Informationstechnik Berlin. – Berlin: ZIB-Report, 2005.
https://www.ifam.uni-hannover.de/fileadmin/ifam/ordner/steinbach/publications/final/14TOAMJ.pdf
17. Pulido–Calvo, I. Gutiérrez–Estrada, J.C.: Selection and operation of pumping stations of water distribution systems. In: Environmental Research Journal, Nova Science Publishers. – 2011, Vol. 5, Issue 3, pp. 1–20. https://sswm.info/node/4007
18. B. Lipták. Pumping station optimization. In: Control Promoting Excellence in Process Automation, pp. 12–19 (2009) https://www.controlglobal.com/manage/optimization/article/11381799/optimization-pumping-station-optimization-part-1-control-global
19. Sergey Dyadun. Information Technologies to Estimation the Effectiveness of Water Supply Systems Control Depending on the Degree of Model Uncertainty // ”ICT in Education, Research, and Industrial Applications: Integration, Harmonization, and Knowledge Transfer” – ICTERI ’2020’ / Kharkiv, V.N. Karazin National University, 2020, pp. 137-145. http://ceur-ws.org/Vol-2740/20200137.pdf https://dblp.uni-trier.de/db/conf/icteri/icteri2020.html
20. Dyadun S.V., Yakovlev S.V., Kobylin O.A. Mathematical Modeling of Steady Flow Distribution in Water Supply Networks with Pumping Stations and Regulating Capacitances // 2nd International Workshop of IT-professionals on Artificial Intelligence (ProfIT AI 2022), 2-4.12.2022, Łódź, Poland. 2022. p. 78-83. https://ceur-ws.org/Vol-3348/ https://ceur-ws.org/Vol-3348/short2.pdf
21. Елизаров Е.Я., Савченко В.С. Численные методы нелинейного программирования. Донецк, 1982, 66с.

22. Койда Н.У., Мархель Э.Г. Расчет оптимального потокораспределения в сети с несколькими точками питания // Тез. докл. XVIII респ. конф. Ровно, 1969, с.21-23.

REFERENCES

1. A.G. Evdokimov, A.D. Tevyashev, *Operational control of the flow distribution in engineering networks*. Vishcha Shkola, Kharkov, 144 p. [in Russian] (1980)
2. A.G. Evdokimov, A.D. Tevyashev, V.V. Dubrovskiy, *Modeling and optimization of the flow distribution in engineering networks.* Stroyizdat, Moscow, 368 p. [in Russian] (1990)
3. A.G. Evdokimov, A.G., Tevyashev, A.D., V.V. Dubrovskiy, *Streaming distribution in engineering networks*. Stroyizdat, Moscow, 199 p. [in Russian] (1979)
4. D. Himmelblau, *Applied nonlinear programming.* Moscow, 534 p. [in Russian] (1975)
5. L.A. Rastrigin, *Extreme control systems.* Moscow, 630 p. [in Russian] (1974)
6. A.G. Evdokimov, *Minimization of functions and its application to the tasks of automatized control of engineering networks.* Kharkov, 288 p. [in Russian] (1985)
7. F. Fallside, P.F. Perry, R.H. Burch, K.C. Marlow, *The Development of Modelling and Simulation Techniques Applied to a Computer - Based - Telecontrol Water Supply System.* In: Computer Simulation of Water Resourses Systems, No.12, pp. 617-639 (1975)
8. N.N. Abramov, Theory and Methodology of Calculation of Water Supply and Distribution Systems. Stroyizdat, Moscow, 288 p. [in Russian] (1985)
9. A.P. Merenkov, V.Y. Hasilev, *Theory of hydraulic circuits.* Nauka, Moscow, 279 p. [in Russian] (1985)
10. Energy Pipeline System: methodological and applied problems of mathematical modeling / N.N. Novitsky, M.G. Suharev, A.D. Tevyashev et al. - Novosibirsk: Science, 476 p. [in Russian] (2015) http://51.isem.irk.ru/semtps/works.php
11. A.D. Tevyashev, O.I. Matvienko, *About one approach to solve the problem of management of the development and operation of centralized water–supply systems.* In: Econtechmod. An International Quarterly Journal. Vol. 3, Issue 3, pp. 61–76 (2014) https://openarchive.nure.ua/server/api/core/bitstreams/d73ad31a-bb66-46ac-987f-14a1943d1b1b/content
12. N.N. Novitskii, O.V. Vanteyeva, *Modeling of stochastic hydraulic conditions of pipeline systems* // Chaotic Modeling and Simulation (CMSIM). No.1, P. 95-108 (2014) http://www.asmda.es/images/Corrected-BOOK_OF_ABSTRACTS-ASMDA2017-5-22.pdf
13. S.V. Dyadun, *Selection of the optimal combinations of the pump station units for the city water supply system.* In: City utilities. Kiev: Tehnika, Issue 1, pp. 63-70 (1992)
14. S.V. Dyadun, *Modeling and rational control of water supply systems with minimum volume of operative information.* In: RadioElectronics & Informatics Journal. Kharkov, KNURE, No. 20, pp. 111-115 [in Russian] (2002) https://cyberleninka.ru/article/n/modelirovanie-i-ratsionalnoe-upravlenie-sistemami-vodosnabzheniya-pri-minimalnom-obeme-operativnoy-informatsii
15. C. Reinbold, V. Hart, *The search for energy savings: optimization of existing & new pumping stations.* In: Florida Water Resources Journal, pp. 44–52 (2011)
16. J. Burgschweiger, B.B. Gnadig, M.C. Steinbach, *Nonlinear programming techniques for operative planning in large drinking water networks.* In: Konrad–Zuse–Zentrum fur Informationstechnik Berlin. Berlin: ZIB-Report. (2005) https://www.ifam.uni-hannover.de/fileadmin/ifam/ordner/steinbach/publications/final/14TOAMJ.pdf
17. I. Pulido–Calvo, J.C. Gutiérrez–Estrada, *Selection and operation of pumping stations of water distribution systems.* In: Environmental Research Journal, Nova Science Publishers. Vol. 5, Issue 3, pp. 1–20 (2011) https://sswm.info/node/4007
18. B. Lipták. *Pumping station optimization.* In: Control Promoting Excellence in Process Automation, pp. 12–19 (2009) https://www.controlglobal.com/manage/optimization/article/11381799/optimization-pumping-station-optimization-part-1-control-global
19. Sergey Dyadun. *Information Technologies to Estimation the Effectiveness of Water Supply Systems Control Depending on the Degree of Model Uncertainty* // "ICT in Education, Research, and Industrial Applications: Integration, Harmonization, and Knowledge Transfer" – ICTERI '2020' /

Kharkiv, V.N.Karazin National University, pp. 137-145. (2020) http://ceur-ws.org/Vol-2740/20200137.pdf
https://dblp.uni-trier.de/db/conf/icteri/icteri2020.html
20. Dyadun S.V., Yakovlev S.V., Kobylin O.A. Mathematical Modeling of Steady Flow Distribution in Water Supply Networks with Pumping Stations and Regulating Capacitances // 2nd International Workshop of IT-professionals on Artificial Intelligence (ProfIT AI 2022), 2-4.12.2022, Łódź, Poland -2022, Scopus , p. 78-83. https://ceur-ws.org/Vol-3348/ https://ceur-ws.org/Vol-3348/short2.pdf
21. Elizarov E.Ya., Savchenko V.S. *Numerical methods of nonlinear programming*. Donetsk. 66 p. [in Russian] (1982)
22. Koida N.U., Markhel E.G. *Calculation of optimal flow distribution in a network with several power points* // XVIII rep. conference. Rivne, p.21-23. [in Russian] (1969)

Dyadun Sergey *PhD on Technical Sciences, Associate Professor;*
Associate Professor of the Department of Modeling Systems and Technologies;
V.N. Karazin Kharkiv National University
4 Svobody Sq., Kharkiv, 61022, Ukraine

Analysis of the effectiveness of flow distribution optimization methods in water supply systems with a large number of pumping stations

Relevance. Currently, optimization methods for a small number of active sources working on the network have been studied. However, when developing operational control systems for water supply systems (WSS) of large cities, one has to deal with a large number of pumping stations (PS) simultaneously working to the network. The complexity of solving the problem of optimization of flow distribution in WSS increases with the increase in the number of active sources working together, which are variables of the optimization of the problem under consideration.
Goal. In the problem of operational control of the modes of operation of the WSS, the task of optimizing flow distribution in a large-scale water supply network occupies an important place. The purpose of the task is to distribute the load (expenditure) between the stations in such a way that, while ensuring the specified quality of water supply to all consumers, the minimum amount of energy consumption at the pumping stations is achieved. The formulation of the problem, the methods of its solution for the WSS of a large city, for which a large number of pumping stations work, are considered. It is necessary to conduct a comparative analysis of the effectiveness of the use of various optimization methods to solve the problem of optimal load distribution among a large number of pumping stations simultaneously working to the water supply system of the metropolis.
Research methods. This problem can be solved by methods of nonlinear mathematical programming or search optimization based on the hydraulic calculation of the water supply network. Its specific feature is the algorithmic task of the goal function. When working to a network of two active sources, this problem is reduced to a problem of one-dimensional search optimization. With a larger number of variables, it is necessary to use methods of multidimensional optimization. The most effective and common methods were used to study the effectiveness of solving the problem of flow distribution optimization in the WSS: coordinate descent; scanning with a variable step; deformed Nelder-Mead polyhedron; Hook and Jeeves direct search; Rosenbrock; Powell.
The results. The conducted research showed that the method of direct search of Hook and Jeeves was the most effective according to the criteria of the minimum expenditure of computer time and the amount of computer memory.
Conclusions. It is advisable to use the obtained results for the development and operation of systems for the operational management of the operation modes of the WSS of large cities, the control systems of dispatchers of water supply networks, CAD of water supply systems to determine the optimal modes of operation of the WSS.

Keywords*: mathematical model, water supply system, pump station, functioning, stream distribution, operational control, criterion, method, efficiency.*

Надійшла у першій редакції 15.03.2024, в останній - 18.04.2024.
The first version has been received on 15.03.2024, the final version - on 18.04.2024.

DOI: https://doi.org/10.26565/2304-6201-2024-61-04

УДК (UDC) 004.12, 004.057.4

Зац Олександр Дмитрович *аспірант факультету комп'ютерних наук*
Харківський національний університет імені В.Н. Каразіна, майдан Свободи, 6, Харків, Україна, 61022
e-mail: zats2021ki51@student.karazin.ua
https://orcid.org/0000-0002-7623-9187

Стрілець Вікторія Євгенівна *к.т.н., доцент кафедри теоретичної та прикладної системотехніки*
Харківський національний університет імені В.Н. Каразіна, майдан Свободи, 6, Харків, Україна, 61022
e-mail: viktoria.strilets@karazin.ua
https://orcid.org/0000-0002-2475-1496

Шматков Сергій Ігорович *д.т.н., проф., завідувач кафедри теоретичної та прикладної системотехніки*
Харківський національний університет імені В.Н. Каразіна, майдан Свободи, 6, Харків, Україна, 61022
e-mail: s.shmatkov@karazin.ua;
https://orcid.org/0000-0002-0298-7174

Ющенко Владислав Сергійович *студент факультету комп'ютерних наук*
Харківський національний університет імені В.Н. Каразіна, майдан Свободи, 6, Харків, Україна, 61022
e-mail: vladyan.yuschenko@gmail.com;
https://orcid.org/0009-0003-8124-4022

Віртуалізація мереж – підхід до оптимізації комп'ютерних мереж

Мета роботи полягає в дослідженні існуючих методів оптимізації комп'ютерних мереж і аналізі підходу віртуалізації мереж як засобу оптимізації. Об'єктом роботи є процес оптимізації комп'ютерних мереж, а предметом – моделі, методи та інформаційні технології, які застосовуються для оптимізації мереж.
Методи дослідження: методи імітаційного і математичного моделювання, методи оптимізації, методи управління, нейромережеві методи.
У **результаті** роботи проведений аналіз підходів і методів оптимізації комп'ютерних мереж. Серед них виділені методи оптимізації топології мереж, методи нелінійної оптимізації параметрів і функціональних залежностей, які описують поведінку і стан мережі. Зазначено, що перспективним є впровадження методів машинного навчання у моделі оптимізації комп'ютерних мереж через їх здатність до узагальнення, класифікації та прогнозування можливих зміни в структурі мережі для покращення її ефективності. Основна увага приділена підходу віртуалізації, який дозволяє абстрагуватися від топології мережі, оптимізувати використання ресурсів, покращити безпеку, спростити керування та забезпечити високий рівень доступності. Такі моделі можуть бути адаптовані до конкретних вимог та обмежень. Серед існуючих напрямків віртуалізації детально розглянуті віртуалізація функцій мереж, побудова програмно-конфігурованих мереж і мережі, визначені знаннями.
Висновки: запропоновано поєднати підхід віртуалізації з методами машинного навчання, а саме побудувати модель оптимізації мережі, яка визначається знаннями, на основі графових нейронних мереж. Такий підхід дасть можливість поєднати складний взаємозв'язок між топологією, маршрутизацією та вхідним трафіком мережі і отримувати точні оцінки розподілу затримок і втрат у мережі.

Ключові слова: *комп'ютерні мережі, оптимізація мереж, управління мережами, віртуалізація, графові нейронні мережі.*

Як цитувати: Зац О. Д., Стрілець В. Є., Шматков С. І., Ющенко В. С. Віртуалізація мереж – підхід до оптимізації комп'ютерних мереж. *Вісник Харківського національного університету імені В.Н. Каразіна, серія Математичне моделювання. Інформаційні технології. Автоматизовані системи управління*. 2024.вип. 61. С.33-43. https://doi.org/10.26565/2304-6201-2024-61-04

How to quote: Zats O.D., Strilets V.Y., Shmatkov S.I., Yuschenko V.S. "Networks virtualization as an approach to optimization of computer networks." *Bulletin of V.N. Karazin Kharkiv National University, series Mathematical modelling. Information technology. Automated control systems*, vol. 61, pp. 33-43, 2024. [In Ukrainian] https://doi.org/10.26565/2304-6201-2024-61-04

1 Вступ

Комп'ютерні мережі стали невід'ємною частиною сучасного світу. Вони використовуються для передачі даних, спільної роботи, комунікації та забезпечення доступу до інтернету. Швидкий

розвиток технологій і зростання обсягів даних, збільшення випадків небажаних втручань призводять до того, що оптимізація комп'ютерних мереж стає більш актуальною та важливою задачею.

Оптимізація відіграє вирішальну роль у точному налаштуванні продуктивності мережі, зменшенні відмов і захисті від кіберзагроз. Оптимізація в комп'ютерних мережах напрямлена на покращення загальної продуктивності, надійності та використання мережевих ресурсів. Це передбачає збалансування різних факторів, таких як пропускна здатність, затримка, пропускна здатність і час відгуку, щоб забезпечити безперебійну передачу даних і взаємодію з користувачем.

Оптимізація комп'ютерних мереж виконується через поєднання оновлень апаратного забезпечення, удосконалення програмного забезпечення та ефективного проектування мереж [1].

Одним із підходів до оптимізації й управління мережами передачі даних є віртуалізація. Сам термін «віртуалізація» означає можливість абстрагування від серверів реальних фізичних компонентів (апаратного забезпечення), і зробити їх придатними для використання у формі віртуальних ресурсів (програмного забезпечення) [2]. Віртуалізація дозволяє обмежити кількість апаратних ресурсів, що значно скорочує витрати на експлуатацію та обслуговування. Повністю відтворюючи фізичну мережу, віртуалізація мережі дозволяє запускати програми у віртуальній мережі, аналогічній фізичній мережі, але з більшими експлуатаційними перевагами та всіма перевагами незалежності від типового апаратного забезпечення віртуалізації [2].

Мета роботи полягає в дослідженні існуючих методів оптимізації комп'ютерних мереж і аналізі сучасного підходу віртуалізації мереж як засобу оптимізації. Об'єктом роботи є процес оптимізації комп'ютерних мереж, а предметом – моделі, методи та інформаційні технології, які застосовуються для оптимізації мереж.

2 Аналіз підходів до оптимізації комп'ютерних мереж

Оптимізація локальної комп'ютерної мережі може включати в себе різні аспекти, включаючи архітектурні, апаратні та програмні рішення.

Одним із підходів є оптимізація топології комп'ютерних мереж [3, 4, 5]. Топологічна модель локальної комп'ютерної мережі (LAN) описує фізичне розташування та з'єднання пристроїв у мережі. Оптимізація топології LAN може включати в себе розгляд різних факторів, таких як продуктивність, надійність, масштабованість та безпека. Вибір оптимальної топології повинен бути підтриманий відповідними налаштуваннями, апаратними рішеннями та методами безпеки.

У роботі [3] для вирішення завдань з оптимізації топології, визначення оптимальних маршрутів, і тому подібне, топологія мережі представлена у вигляді графу G=(V,E), де множина вершин V відповідає вузлам мережі, а множина ребер E – каналам. Застосування теорії графів надає можливість провести детальне дослідження мережі, охоплюючи характеристики ребер графу, за допомогою яких можна аналізувати та впливати на такі поняття як: пропускна здатність, завантаженість, чи затримки. Як тільки граф буде відповідати топології комп'ютерної мережі, використовують метод аналізу ієрархій, для дослідження ефективності фізичної структури, і чи здатна вона забезпечувати задовільний рівень якості надання послуг.

Іншим підходом до оптимізації LAN є створення і дослідження математичних моделей LAN [4, 6]. Математична модель оптимізації локальної комп'ютерної мережі може бути досить складною і залежить від конкретних цілей та обмежень. Основна мета такої моделі – мінімізувати певний параметр чи функцію, такі як: вартість, час затримки, або споживану енергію, з урахуванням різних змінних і обмежень мережі. У роботі [4] запропонована модель проектування оптимізації мережі, заснована на встановленні максимізації надійності мережі за заданих обмежень вартості. А у роботі [6] розглянута модель багатокритеріальної оптимізації на комбінаторній конфігурації вузлів, яка дає можливість оптимізувати характеристики роботи комп'ютерних мереж багаторівневої архітектури.

Підходи із використанням моделей машинного навчання для оптимізації LAN [7, 8] можуть бути корисними для автоматизації процесів моніторингу та управління мережею з метою підвищення її продуктивності та надійності. У роботі [7] узагальнені підходи і технології машинного навчання, які використовуються дослідниками для аналізу, управління, моніторингу і оптимізації мереж. У роботі [8] були розглянуті методи машинного навчання для підвищення продуктивності та масштабованості алгоритмів маршрутизації.

З аналізу літературних джерел випливає, що існують і використовуються різні підходи до управління і оптимізації комп'ютерних мереж, які вирішують різні задачі оптимізації: підвищення

ефективності мереж через удосконалення топології, оптимізація інфраструктури мереж, оптимізація маршрутизації та ін. Але ці підходи не мають можливості відтворити і дослідити фізичну модель мережі, тому пропонується розглянути підхід віртуалізації та його застосування до оптимізації мереж.

3 Поняття віртуалізації мереж

Віртуалізація локальної комп'ютерної мережі (LAN) [9] може бути важливим кроком для оптимізації та управління мережею у великих організаціях або компаніях, де важливо забезпечити високий рівень ефективності, безпеки та доступності. Віртуалізаційна модель LAN складається з:

– віртуалізації комутації (Network Virtualization). Використання віртуальних інтерфейсів на комутаторах, які можуть бути програмно налаштовані та управлятися. Це дозволяє легко масштабувати мережу та налаштовувати маршрутизацію для оптимізації трафіку. Розділення LAN на віртуальні мережі (VLAN) ізолює різні групи пристроїв та зменшить затори.

– віртуалізації ресурсів (Resource Virtualization). Використання віртуальних серверів та обчислювальних ресурсів для оптимізації використання обладнання. Впровадження віртуальних машин дозволяє ефективніше розподіляти обчислювальні завдання та зменшити витрати на обладнання.

– віртуалізації безпеки (Security Virtualization). Можна використовувати віртуальні файєрволи, IDS/IPS системи та інші засоби безпеки для захисту мережі та даних. При цьому доступ до ресурсів мережі може бути керованим та спостерігатися на рівні віртуальних мереж та сегментів.

– віртуалізації моніторингу (Monitoring Virtualization). Включає в себе використання засобів моніторингу мережі, які дозволяють відстежувати стан та продуктивність віртуальних ресурсів. Автоматизована система моніторингу дозволяє вчасно виявляти проблеми та вживати заходи для їх вирішення.

– віртуалізації керування (Management Virtualization). Використання централізованого програмного забезпечення для керування всією віртуальною LAN. При цьому забезпечується можливість автоматизованого розгортання, конфігурації та моніторингу віртуальних ресурсів.

Віртуалізація LAN дозволяє оптимізувати використання ресурсів, покращити безпеку, спростити керування та забезпечити високий рівень доступності. Така модель може бути адаптована до конкретних вимог та обставин вашої організації.

4 Методологія віртуалізації мереж

Віртуалізація мереж виконується у двох формах: зовнішня і внутрішня. Обидві форми відповідають їх розташуванню по відношенню до сервера. Зовнішня віртуалізація використовує комутатори, адаптери або мережі для об’єднання однієї чи кількох мереж у віртуальні одиниці. Внутрішня віртуалізація використовує мережеві функції в програмних контейнерах на одному мережевому сервері, що дозволяє віртуальним машинам обмінюватися даними на хості без використання зовнішньої мережі [9].

Вид віртуалізації мережі зазвичай визначається за їх використанням у різних сегментах мережі, таких як центр обробки даних, WAN або LAN. Програмно-визначена мережа (SDN) привела до еволюції віртуалізації мережі центрів обробки даних, а поява програмно-визначеної глобальної мережі (SD-WAN) зробила революцію у віртуалізації WAN. Тоді як віртуалізація локальної мережі LAN необхідна підприємствами, які впроваджують програмно визначену локальну мережу (SD-LAN) для покращення операцій.

Розглянемо різні напрямки віртуалізації, які можуть бути корисними для оптимізації мереж.

4.1 Віртуалізація функцій мережі

Віртуалізація функцій мережі [10] (network function virtualization – NFV) є концепцією мережевої архітектури, в якій за допомогою віртуалізації замінюються функції мережевих пристроїв. Технологія віртуалізації мережевих функцій поєднує функції таких мережевих пристроїв, як брандмауери, балансувальники навантаження і аналізатори трафіку, що працюють разом для підвищення продуктивності мережі.

Архітектура NFV складається з трьох частин:

– централізована інфраструктура віртуальної мережі (NFVI): інфраструктура NFV може базуватися або на платформі керування контейнером, або на гіпервізорі, який абстрагує сховище, обчислювальні, та мережеві ресурси.

– програмні додатки або функції віртуалізованої мережі (VNF): програмне забезпечення замінює апаратні компоненти традиційної мережевої архітектури для надання різних типів мережевих функцій (віртуалізованих мережевих функцій).

– фреймворк (часто відомий як MANO – управління, автоматизація та мережева оркестровка) необхідний для керування інфраструктурою та надання мережевих функцій.

Переваги NFV:

– дозволяє гнучко та динамічно надавати нові послуги, скоротивши при цьому капітальні та операційні витрати;

– заміна обладнання на стандартизовані сервери та мережеві компоненти не прив’язує операторів до постачальників обладнання;

– знижує операційні витрати за рахунок спрощення моніторингу та адміністрування операцій (всі мережеві функції переносяться в єдину віртуалізовану інфраструктуру);

– швидке підключення нових користувачів до мережі;

– окупає інфраструктуру телекомунікаційних компаній.

На рис. 1 зображено порівняння традиційної мережі із NFV мережею.

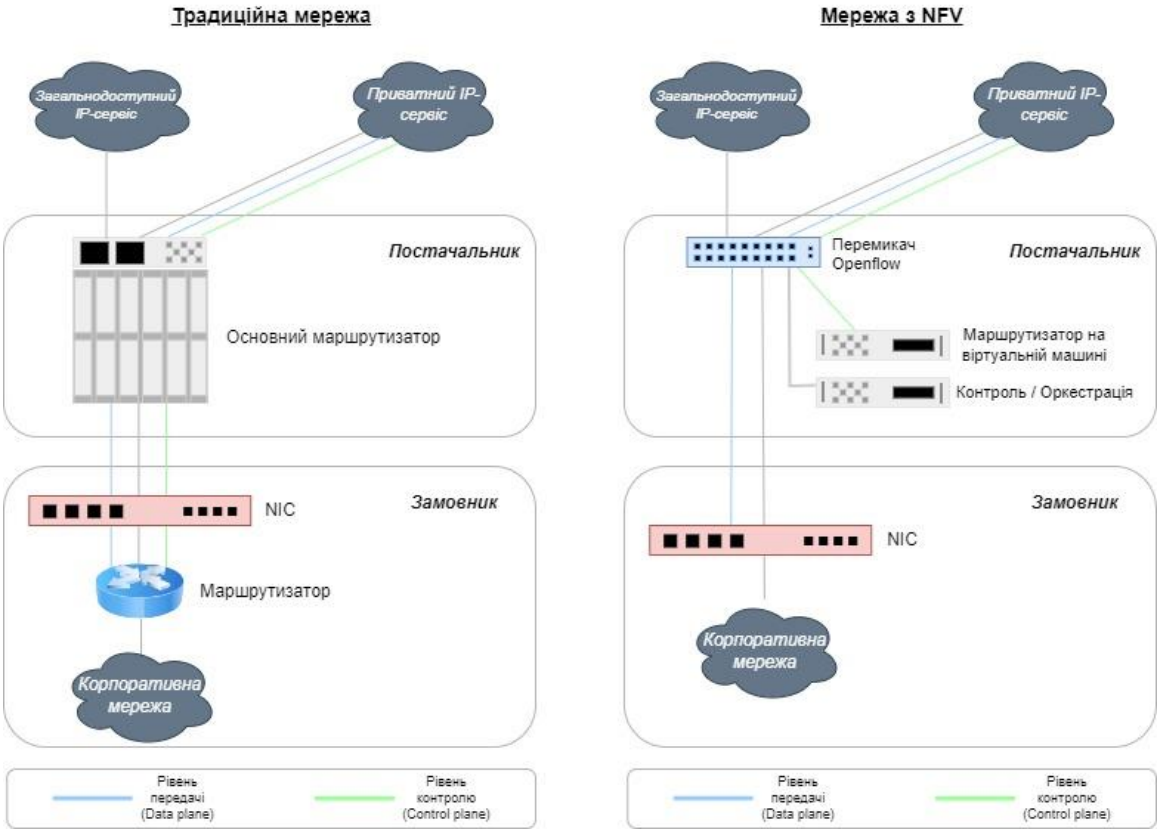


Рисунок 1. Порівняння традиційної мережі із NFV мережею

Основна відмінність мережі NFV полягає в тому, що на стороні сервіс провайдеру, знаходиться не один маршрутизатор, а віртуальні та фізичні пристрої, які можуть динамічно визначати шлях призначення пакету, тому на стороні користувача відсутні маршрутизатори, адже пакети надходять напряму до кінцевих пристроїв.

4.2 Програмно-конфігурована мережа

Програмно-конфігурована мережа (Software-defined Networking – SDN) – це підхід до мереж, який використовує програмні контролери, що можуть керуватися інтерфейсами прикладного програмування (API) для зв’язку з апаратною інфраструктурою для спрямування мережевого

трафіку. Використовуючи програмне забезпечення, можна створити та керувати низкою віртуальних накладених мереж, які працюють у поєднанні з фізичною базовою мережею. Мережі SDN пропонують потенціал для доставки середовищ програмних застосунків у вигляді коду та мінімізації практичного часу, необхідного для керування мережею [12].

У SDN програмне забезпечення відокремлено від апаратного забезпечення. Мережа SDN складається із двох площин: площини керування та площини даних. Площина керування визначає куди надсилати трафік програмному забезпеченню, в той час коли площина даних фактично пересилає трафік в апаратне забезпечення. Це дозволяє мережевим адміністраторам програмувати та керувати всіє мережею з однієї точки, а не від пристрою до пристрою.

Типова архітектура SDN (рис. 2) складається із трьох частин, які можуть бути розташовані в різних фізичних місцях, та двох інтерфейсів.

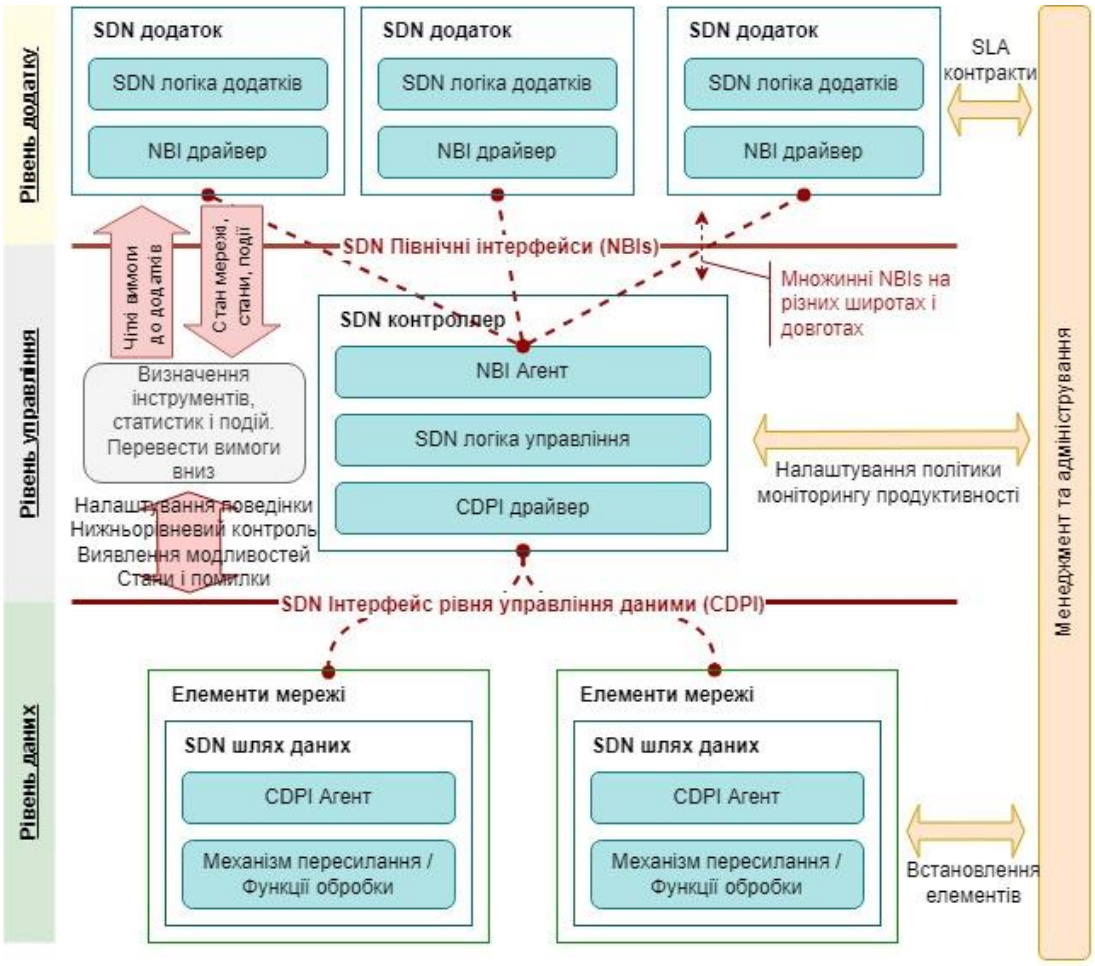


Рисунок 2. Загальна архітектура SDN

Програми SDN – це програми, які явно, прямо та програмно передають свої вимоги до мережі та бажаної мережевої поведінки контролеру SDN через північний інтерфейс (NBI). Крім того, вони можуть використовувати абстрактне уявлення про мережу для своїх внутрішніх цілей прийняття рішень. Програма SDN складається з однієї логіки програми SDN і одного або кількох драйверів NBI. Програми SDN можуть самі надавати інший рівень абстрактного мережевого контролю, таким чином пропонуючи один або більше NBI вищого рівня через відповідних агентів NBI.

Контролер SDN – це логічно-централізований об’єкт, який відповідає за переклад вимог із прикладного рівня SDN до шляхів даних SDN і надання додаткам SDN абстрактного представлення мережі (яке може включати статистику та події) . Контролер SDN складається з одного або кількох агентів NBI, логіки керування SDN і драйвера інтерфейсу керування до площини даних (CDPI). Хоча визначення, як логічно-централізованого об’єкта не передбачає і не виключає таких деталей реалізації як: об’єднання кількох контролерів; ієрархічне з’єднання

контролерів; інтерфейси зв'язку між контролерами, а також віртуалізація чи нарізка мережевих ресурсів.

Мережеві пристрої отримують інформацію від контролерів про те, куди перемістити дані.

Інтерфейс Control to Data-Plane Interface (CDPI) – це інтерфейс, визначений між контролером і мережевими пристроями, який забезпечує: програмне керування всіма операціями пересилання, оголошення про можливості, статистичні звіти та повідомлення про події.

Північний інтерфейс (NBI) – це інтерфейс, визначений між додатками і контролерами, які зазвичай надають абстрактні мережеві представлення та забезпечують пряме вираження мережевої поведінки та вимог.

Фізичні або віртуальні мережеві пристрої фактично переміщують дані через мережу. У деяких випадках віртуальні комутатори, які можуть бути вбудовані в програмне або апаратне забезпечення, беруть на себе обов'язки фізичних комутаторів і об'єднують їхні функції в єдиний інтелектуальний комутатор. Комутатор перевіряє цілісність як пакетів даних, так і місць призначення віртуальної машини та переміщує пакети.

Існують різні моделі SDN:

– Open SDN: мережеві адміністратори використовують такий протокол, як OpenFlow, щоб керувати поведінкою віртуальних і фізичних комутаторів на рівні даних.

– SDN через API: Замість використання відкритого протоколу інтерфейси прикладного програмування контролюють, як дані переміщуються через мережу на кожному пристрої.

– Модель SDN Overlay. Інший тип програмно визначеної мережі запускає віртуальну мережу поверх існуючої апаратної інфраструктури, створюючи динамічні тунелі до різних локальних і віддалених центрів обробки даних. Віртуальна мережа розподіляє пропускну здатність для різних каналів і призначає пристрої для кожного каналу, залишаючи фізичну мережу недоторканою.

– Hybrid SDN: ця модель поєднує програмно визначену мережу з традиційними мережевими протоколами в одному середовищі для підтримки різних функцій у мережі. Стандартні мережеві протоколи продовжують спрямовувати частину трафіку, тоді як SDN бере на себе відповідальність за інший трафік, дозволяючи мережевим адміністраторам поетапно вводити SDN у застаріле середовище.

Ключова відмінність між SDN і традиційною мережею полягає в інфраструктурі: SDN базується на програмному забезпеченні, тоді як традиційна мережа базується на апаратному забезпеченні. Оскільки площина керування базується на програмному забезпеченні, SDN є набагато гнучкішим, ніж традиційна мережа. Це дозволяє адміністраторам контролювати мережу, змінювати параметри конфігурації, надавати ресурси та збільшувати пропускну здатність мережі – і все це через централізований інтерфейс користувача без додавання додаткового обладнання.

Існують також відмінності в безпеці між SDN і традиційними мережами. Завдяки більшій видимості та можливості визначати безпечні шляхи, SDN пропонує кращу безпеку багатьма способами. Однак, оскільки програмно визначені мережі використовують централізований контролер, безпека контролера має вирішальне значення для підтримки безпечної мережі, і це єдиний вузол що представляє потенційну вразливість SDN.

4.3 Мережа, визначена знаннями

Мережа, визначена знаннями (Knowledge-Defined Networking – KDN) [13] – це розширена версія SDN, яка робить крок вперед, відокремлюючи площину керування від логіки керування та вводячи нову площину, яка називається площиною знань, відокремленою від логіки керування для генерування знань на основі даних, зібраних із мережі.

Мережа, визначена знаннями (KDN) – це концепція використання інформації для генерування знань за допомогою моделей машинного навчання або моделей на основі правил, і відповідно до цих знань приймаються мережеві рішення.

KDN складається із п'яти основних площин. Блок-схема високого рівня архітектури показана на рис. 3.

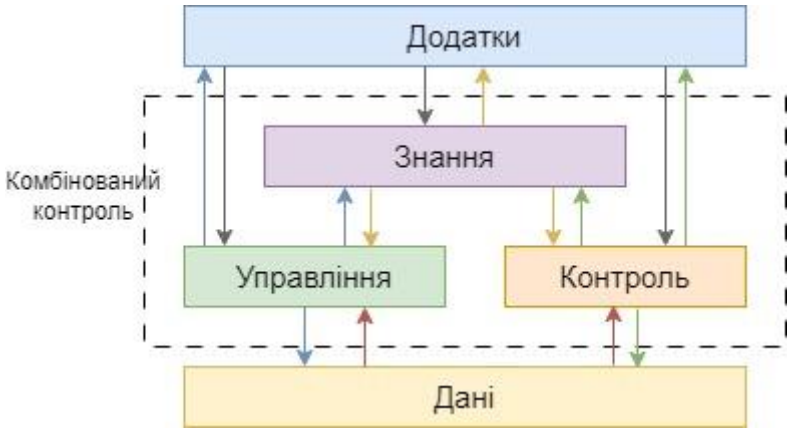


Рисунок 3. Схема високого рівня архітектури KDN

Площина знань складається з трьох підрівнів:

– площина генерації знань – генерує знання з використанням даних/інформації за допомогою методів на основі евристичної моделі або методів машинного навчання;

– площина композиції знань – компонує згенеровані знання та універсальні знання, за допомогою редактора онтології, для створення складених знань, які можна використовувати для створення правил шляхом узгодження з намірами користувача;

– площина розподілу та управління знаннями – зберігає знання та правила з використанням бази знань.

Площина управління працює паралельно з контролером KDN і відповідає за: збір процесів і даних/інформації з мережевих пристроїв, моніторинг стану мережевого пристрою та налаштування мережевого пристрою.

Площина даних складається з пристроїв пересилання, які можуть зберігати, пересилати або обробляти дані відповідно до правил потоку, надісланих площиною керування. У KDN площина даних потрібна для надсилання даних, запитуваних площинами управління та контролю.

Площина контролю складається з одного або декількох контролерів SDN на основі архітектури та відповідає за надсилання правил потоку, правил контролю доступу, правил пріоритизації трафіку на основі QoS тощо до площини даних.

Площина додатків забезпечує платформу для мережевих додатків для передачі вимог базовій мережевій інфраструктурі. Це також дозволяє мережевим адміністраторам централізовано визначати мережеві політики, специфічні для додатків, і визначати політики конфігурації мережі, які більш узгоджені з бізнес-потребами та цілями високого рівня, де логіка програми відокремлена від апаратного забезпечення.

5 Оптимізація мережі з використанням віртуалізації

Розглянемо підхід до оптимізації комп’ютерної мережі в контексті парадигми мережі, визначеною знаннями (KDN). У такому випадку припускається, що площина керування отримує постійні оновлення стану мережі (наприклад, дані трафіку, показники затримки). Показники стану мережі можна формувати за допомогою «звичайних» методів вимірювання на основі SDN. У площині знань є оптимізатор, який визначається заданою цільовою функцією (політикою) (рис. 4). Ця політика, відповідно до мереж на основі знань, може бути визначена декларативною мовою, наприклад NEMO [14], і представлена у вигляді багатоцільової задачі оптимізації мережі.

Точна мережева модель може мати головну роль під час оптимізації. Використовують її для ініціалізації алгоритмів, які ітеративно визначають ефективність побудованих рішень, щоб знайти найкращу конфігурацію мережу. За межами архітектури мережі, визначеної знаннями, навмисно залишають етап навчання. В таких випадках мережева модель повинна відповідати двом основним вимогам: забезпечувати точність результатів та мати низьку обчислювальну складність, що дозволить оптимізаторам мережі знаходити рішення за короткі проміжки часу. Крім того, оптимізаторам важливо мати достатню гнучкість для моделювання сценаріїв «якщо-то», які можуть враховувати різні схеми маршрутизації, зміни в топології та варіації в даних трафіку.

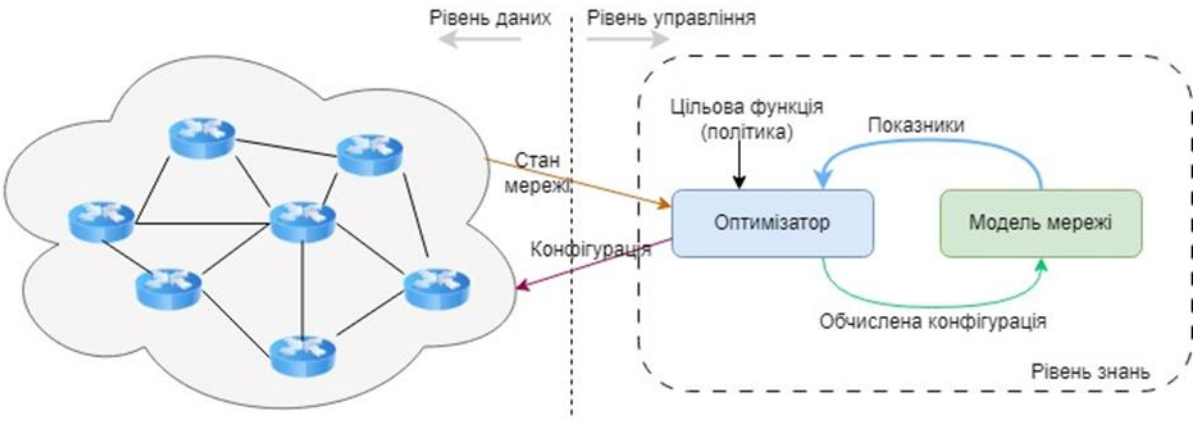


Рисунок 4. Модель оптимізація мережі на основі KDN

У таких випадках для створення моделей мереж, визначених знаннями, звертаються до машинного навчання, а саме використовують графові нейронні мережі (Graph Neural Network, GNN) [15]. Моделі GNN здатні ефективно працювати та узагальнювати середовища, представлені у вигляді графів.

У роботі [16] представлена графова нейронна мережа RouteNet, побудована на основі нейронних мереж передачі повідомлень, здатна поширювати будь-яку схему маршрутизації по топології мережі та абстрагувати значущу інформацію про поточний стан мережі. На рис. 5 показано схематичне зображення моделі нейронної мережі RouteNet.

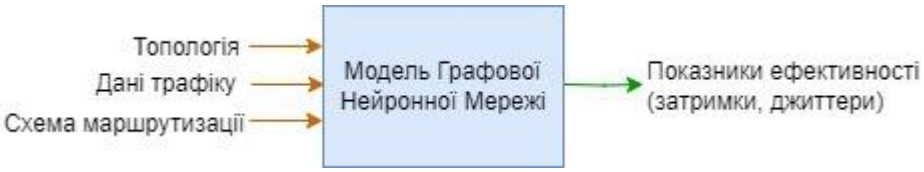


Рисунок 5. Схема RouteNet

RouteNet приймає як вхідні дані задану топологію, схему маршрутизації, джерело-призначення (тобто зв'язки між кінцевими точками, шляхами та посиланнями) і матрицю трафіку (визначену як пропускна здатність між кожною парою вузлів у мережі). І на виході надає показники продуктивності відповідно до поточного стану мережі (наприклад, затримки на шлях або тремтіння сигналу). Для цього RouteNet використовує вектори з фіксованою розмірністю, які кодують стани шляхів і посилань і передають інформацію між ними відповідно до схеми маршрутизації.

Таким чином, для подальших досліджень і побудови моделей оптимізації комп'ютерних мереж були обрані графові нейронні мережі, які здатні розуміти складний взаємозв'язок між топологією, маршрутизацією та вхідним трафіком мережі і отримувати точні оцінки розподілу затримок і втрат для кожного джерела/адресата.

6 Висновки

Розглянувши підходи віртуалізації мереж можна зробити висновок, що NFV, SDN та KDN є важливими інноваціями, які значно підвищують ефективність мережевих інфраструктур порівняно з традиційними методами.

Традиційна мережа є найстарішим підходом до роботи в мережі та передбачає ручне налаштування та керування пристроями. Цей метод мережевих зв'язків був поширеним із самого початку створення мереж і досі переважає в сучасних мережах зв'язку.

SDN — це новітній підхід, який відокремлює площину керування від площини даних і забезпечує більшу гнучкість у проектуванні мережі. Площина керування переміщується в централізоване розташування, а мережеві адміністратори використовують програмне забезпечення для керування мережею. Ця парадигма дозволяє мережевим адміністраторам легше та швидше керувати мережами завдяки підвищеній гнучкості та можливості програмування.

KDN створює мережу, що самонавчається, самооптимізується та самовідновлюється, інтегруючи технології AI та ML. Аналіз даних — це інструмент, який використовується системами KDN для автоматичного покращення продуктивності мережі, адаптації до мінливих мережевих обставин, а також виявлення та вирішення потенційних мережевих проблем до того, як вони стануть серйознішими.

NFV, SDN та KDN мають перевагу над традиційними методами в гнучкості, масштабованості, в швидкості запровадженні змін, в економії ресурсів, легке керування та можливість автоматизації.

А поєднання методів віртуалізації та нейронних мереж створює потужний інструментарій для оптимізації локальної комп'ютерної мережі. Це дозволить досягти балансу між гнучкістю управління та точністю прогнозування значень параметрів мережі, що в свою чергу призводить до підвищення продуктивності та ефективності використання ресурсів.

СПИСОК ЛІТЕРАТУРИ

1. Bashar, Mesfer. Optimization in Computer Networks and Cybersecurity: Ensuring Efficiency and Safety. *Global J Technol Optim*, 14 (2023): 333. DOI: 10.37421/2229-8711.2023.14.333.
2. Network Virtualization: Optimization and Reliability: website. URL: https://www.tnsolutions.it/en/network-virtualization-optimization-and-reliability (дата звернення 1.06.2024)
3. Б. А. Бугиль, О. А. Лаврів, М. І. Бешлей, В. В. Червенець Методи оптимізації фізичної та логічної структур телекомунікаційних мереж. *Вісник Національного університету "Львівська політехніка". Радіоелектроніка та телекомунікації*. 2013. № 766. С. 78-83. URL: https://science.lpnu.ua/sites/default/files/journal-paper/2017/jun/5107/12bugillavrivbeshleychervenec.pdf (дата звернення: 01.06.2024)
4. Ai, Hua, Fan, Yuhong, Zhang, Jilei and Ghafoor, Kayhan Zrar. Topology optimization of computer communication network based on improved genetic algorithm. *Journal of Intelligent Systems*, vol. 31, no. 1, 2022, pp. 651-659. https://doi.org/10.1515/jisys-2022-0050
5. Hadi Rezazad. Computer network optimization. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2011. 3(1), pp. 34 – 46. DOI: 10.1002/wics.135
6. Л.М. Колєчкіна, А.М. Нагірна. Математична модель багатокритеріальної оптимізації на множині сполучень при побудові комп'ютерних мереж. *Математичні машини і системи*. 2016. № 4. С. 68-75.
7. Haitham Afifi, Sabrina Pochaba, Andreas Boltres, Dominic Laniewski and others. Machine Learning with Computer Networks: Techniques, Datasets and Models. *IEEE Access*, 2024. 52 p. DOI: 10.1109/ACCESS.2024.3384460.
8. Ke Liang, Mitchel Myers. Machine Learning Application in the Routing in Computer Networks. *ArXiv abs/2104.01946*, 2021. URL: https://arxiv.org/pdf/2104.01946 (дата звернення: 25.05.2024)
9. What is network virtualization? Everything you need to know: website. URL: https://www.techtarget.com/searchnetworking/What-is-network-virtualization-Everything-you-need-to-know (дата звернення: 25.05.2024)
10. Network Functions Virtualization – Introductory White Paper. *SDN and OpenFlow World Congress*, 2012. URL: https://portal.etsi.org/NFV/NFV_White_Paper.pdf (дата звернення: 25.05.2024)
11. В.В. Палагін, І.О. Євтушенко, О.О. Гожий. Віртуалізаця як середовище реалізації мережевих функцій. *Вісник Черкаського державного технологічного університету*, 2021. №2. С. 31-38. DOI: 10.24025/2306-4412.2.2021.234703.
12. What is Software-Defined Networking? IBM : website. URL: https://www.ibm.com/topics/sdn (дата звернення: 10.05.2024).
13. Comprehensive Survey on Knowledge-Defined Networking. MDPI : website. URL: https://www.mdpi.com/2673-4001/4/3/25 (дата звернення: 10.05.2024).
14. NeMo: an application's interface to intent-based networks : website. URL: http://nemo-project.net/ (дата звернення: 15.05.2024).

15. Krzysztof Rusek, José Suárez-Varela, Albert Mestres, Pere Barlet-Ros, and Albert Cabellos-Aparicio. Unveiling the potential of Graph Neural Networks for network modeling and optimization in SDN. *In Proceedings of the 2019 ACM Symposium on SDN Research (SOSR '19)*. Association for Computing Machinery, New York, USA, 2019. Pp. 140–151. DOI: https://doi.org/10.1145/3314148.3314357
16. K. Rusek, J. Suárez-Varela, P. Almasan, P. Barlet-Ros and A. Cabellos-Aparicio. RouteNet: Leveraging Graph Neural Networks for Network Modeling and Optimization in SDN. *IEEE Journal on Selected Areas in Communications*. 2020. V. 38, № 10. P. 2260–2270. DOI: https://doi.org/10.1109/jsac.2020.3000405.

REFERENCES

1. Bashar, Mesfer. Optimization in Computer Networks and Cybersecurity: Ensuring Efficiency and Safety. *Global J Technol Optim*, 14 (2023): 333. DOI: 10.37421/2229-8711.2023.14.333.
2. Network Virtualization: Optimization and Reliability: website. URL: https://www.tnsolutions.it/en/network-virtualization-optimization-and-reliability (дата звернення 1.06.2024)
3. Buhyl B.A., Lavriv O.A., Beshley M.I., Chervenets V.V. Optimization methods for telecommunications networks physical and logical structures. *Bulletin of Lviv Polytechnic. Series of Radio Electronics and Telecommunication. 2013. № 766. Pp. 78-83*. URL: https://science.lpnu.ua/sites/default/files/journal-paper/2017/jun/5107/12bugillavrivbeshleychervenec.pdf (дата звернення: 01.06.2024)
4. Ai, Hua, Fan, Yuhong, Zhang, Jilei and Ghafoor, Kayhan Zrar. Topology optimization of computer communication network based on improved genetic algorithm. *Journal of Intelligent Systems, vol. 31, no. 1, 2022, pp. 651-659*. https://doi.org/10.1515/jisys-2022-0050
5. Hadi Rezazad. Computer network optimization. *Wiley Interdisciplinary Reviews: Computational Statistics, 2011. 3(1), pp. 34 – 46*. DOI: 10.1002/wics.135
6. Koliechkina L.M., Nahirna A.M. A mathematical model of multi-criteria optimization on the set of combinations under the construction of computer networks. *Mathematical machines and systems. 2016. № 4. Pp. 68-75*.
7. Haitham Afifi, Sabrina Pochaba, Andreas Boltres, Dominic Laniewski and others. Machine Learning with Computer Networks: Techniques, Datasets and Models. *IEEE Access, 2024. 52 p*. DOI: 10.1109/ACCESS.2024.3384460.
8. Ke Liang, Mitchel Myers. Machine Learning Application in the Routing in Computer Networks. *ArXiv abs/2104.01946, 2021*. URL: https://arxiv.org/pdf/2104.01946 (дата звернення: 25.05.2024)
9. What is network virtualization? Everything you need to know: website. URL: https://www.techtarget.com/searchnetworking/What-is-network-virtualization-Everything-you-need-to-know (дата звернення: 25.05.2024)
10. Network Functions Virtualization – Introductory White Paper. *SDN and OpenFlow World Congress, 2012*. URL: https://portal.etsi.org/NFV/NFV_White_Paper.pdf (дата звернення: 25.05.2024)
11. Palahin V.V., Yevtushenko I.O., Hozhyi O.O. Virtualization as an environment of realization of network functions. *Bulletin of Cherkasy State Technological University*, 2021. № 2. Pp. 31-38. DOI: 10.24025/2306-4412.2.2021.234703.
12. What is Software-Defined Networking? IBM : website. URL: https://www.ibm.com/topics/sdn (дата звернення: 10.05.2024).
13. Comprehensive Survey on Knowledge-Defined Networking. MDPI : website. URL: https://www.mdpi.com/2673-4001/4/3/25 (дата звернення: 10.05.2024).
14. NeMo: an application's interface to intent-based networks : website. URL: http://nemo-project.net/ (дата звернення: 15.05.2024).

15. Krzysztof Rusek, José Suárez-Varela, Albert Mestres, Pere Barlet-Ros, and Albert Cabellos-Aparicio. Unveiling the potential of Graph Neural Networks for network modeling and optimization in SDN. *In Proceedings of the 2019 ACM Symposium on SDN Research (SOSR '19). Association for Computing Machinery, New York, USA, 2019. Pp. 140–151*. DOI: https://doi.org/10.1145/3314148.3314357
16. K. Rusek, J. Suárez-Varela, P. Almasan, P. Barlet-Ros and A. Cabellos-Aparicio. RouteNet: Leveraging Graph Neural Networks for Network Modeling and Optimization in SDN. *IEEE Journal on Selected Areas in Communications. 2020. V. 38, № 10. P. 2260–2270*. DOI: https://doi.org/10.1109/jsac.2020.3000405.

Zats Oleksandr *PhD student of Computer Science Faculty;*
V.N. Karazin Kharkiv National University, Svobody Sq 6, Kharkiv, Ukraine, 61022
Strilets Viktoriia *Ph.D, associate professor of the theoretical and applied system engineering department;*
V.N. Karazin Kharkiv National University, Svobody Sq 6, Kharkiv, Ukraine, 61022
Shmatkov Serhiy *Doctor of Engineering Sciences, professor, Head of Theoretical and Applied Systems Engineering Department;*
V.N. Karazin Kharkiv National University, Svobody Sq 6, Kharkiv, Ukraine, 61022
Yuschenko Vladyslav *student of Computer Science Faculty;*
V.N. Karazin Kharkiv National University, Svobody Sq 6, Kharkiv, Ukraine, 61022

Networks virtualization as an approach to optimization of computer networks

The **purpose** of the work is to study the existing methods of optimizing computer networks and analyze the approach of virtualization of networks as a means of optimization. The object of the work is the process of optimizing computer networks, and the subject is models, methods and information technologies that are used to optimize networks.
Research methods: simulation and mathematical modeling methods, optimization methods, control methods, neural network methods.
As a **result** of the work, an analysis of computer network optimization approaches and methods was carried out. Among them, optimization methods of network topology, methods of nonlinear optimization of parameters and functional dependencies, which describe the behavior and state of the network, are highlighted. It is noted that the implementation of machine learning methods in the optimization model of computer networks is promising due to their ability to generalize, classify and predict possible changes in the network structure to improve its efficiency. The focus is on a virtualization approach that allows you to abstract from network topology, optimize resource usage, improve security, simplify management, and ensure high availability. Such models can be adapted to specific requirements and constraints. Among the existing directions of virtualization, the virtualization of network functions, the construction of software-configured networks and knowledge-defined networks are considered in detail.
Conclusions: it is proposed to combine the virtualization approach with machine learning methods, namely to build a knowledge-based network optimization model based on graph neural networks. This approach will make it possible to combine the complex relationship between topology, routing and incoming network traffic and obtain accurate estimates of the distribution of delays and losses in the network.

Keywords: *computer networks, network optimization, network management, virtualization, graph neural networks*.

Надійшла у першій редакції 02.03.2024, в останній - 11.04.2024.
The first version has been received on 02.03.2024, the final version - on 11.04.2024.

DOI: https://doi.org/10.26565/2304-6201-2024-61-05

УДК (UDC) 519.688:004.934

Ivaniuk Andrii
PhD Student
National University of "Kyiv-Mohyla Academy", Faculty of Computer Sciences, 2 Skovorody st., Kyiv, Ukraine, 04655
e-mail: a.ivaniuk@ukma.edu.ua
https://orcid.org/0000-0002-4189-3787

Latent diffusion model for speech signal processing

Topicality. The development of generative models for audio synthesis, including text-to-speech (TTS), text-to-music, and text-to-audio applications, largely depends on their ability to handle complex and varied input data. This paper centers on latent diffusion modeling, a versatile approach that leverages stochastic processes to generate high-quality audio outputs.
Key goals. This study aims to evaluate the efficacy of latent diffusion modeling for TTS synthesis on the EmoV-DB dataset, which features multi-speaker recordings across five emotional states, and to contrast it with other generative techniques.
Research methods. We applied latent diffusion modeling to TTS synthesis specifically and evaluated its performance using metrics that assess intelligibility, speaker similarity, and emotion preservation in the generated audio signal.
Results. The study reveals that while the proposed model demonstrates decent efficiency in maintaining speaker characteristics, it is outperformed by the discrete autoregressive model: xTTS v2 in all assessed metrics. Notably, the researched model exhibits deficiencies in emotional classification accuracy, suggesting potential misalignment between the emotional intents encoded by the embeddings and those expressed in the speech output.
Conclusions. The findings suggest that further refinement of the encoder's ability to process and integrate emotional data could enhance the performance of the latent diffusion model. Future research should focus on optimizing the balance between speaker and emotion characteristics in TTS models to achieve a more holistic and effective synthesis of human-like speech.

Keywords: *audio modeling, artificial neural networks, speech synthesis.*

Як цитувати: Ivaniuk A. Latent diffusion model for speech signal processing. *Вісник Харківського національного університету імені В.Н. Каразіна, сер. «Математичне моделювання. Інформаційні технології. Автоматизовані системи управління»*. 2024. вип. 61. С.43-52. https://doi.org/10.26565/2304-6201-2024-61-05

How to quote: Ivaniuk A., "Latent diffusion model for speech signal processing." *Bulletin of V.N. Karazin Kharkiv National University, series Mathematical modelling. Information technology. Automated control systems*, vol. 61, pp. 43-52, 2024. https://doi.org/10.26565/2304-6201-2024-61-05

1. Introduction

The pursuit of sophisticated generative models has invariably involved the integration of robust frameworks that underpin the generation process. At the heart of our study is the development of a model designed to cater specifically to the complexities inherent in generating realistic and nuanced audio outputs. This model is not only pivotal for understanding the theoretical underpinnings of audio synthesis but also serves as the backbone for practical applications in various audio generation tasks, including Text-to-Speech (TTS) systems.

TTS technology, which converts text into spoken voice output, has seen significant advancements through the adoption of deep learning models that improve naturalness and intelligibility. To enhance our model's capabilities within the TTS domain, we align our objectives with those of existing implementations that leverage similar neural architectures. Notably, the implementation of the Audio Latent Diffusion Model 2 (Audio LDM2) provides a basis for its innovative approach to audio synthesis. This model, known for its effectiveness in handling high-dimensional audio data through a diffusion-based process, aligns closely with our goals.

The existence of such a model as Audio LDM2 offers an opportunity to not only refine our approach by tuning our model based on this established framework but also to rigorously compare its performance against current competitive models like xTTS v2. This comparative analysis aims to highlight the limitations and potential our approach brings to the TTS research field, potentially setting new benchmarks for audio quality.

2. Related Work

Text-to-speech (TTS) synthesis has seen significant advancements due to the adoption of deep learning techniques, which have greatly improved the naturalness and expressiveness of synthesized speech. This section reviews several key methodologies in TTS that share, particularly focusing on models that integrate advanced neural network architectures and embeddings to enhance speech quality and emotional expressivity.

Tacotron models. One of the foundational models in modern TTS is Tacotron, which uses a sequence-to-sequence framework with attention to convert text directly into speech [1]. This model laid the groundwork for further developments in end-to-end speech synthesis. Following Tacotron, Tacotron 2 integrated WaveNet, a deep generative model of raw audio waveforms, to improve the naturalness of the speech output [2].

Embedding-Based Models. Significant similarities can be drawn with models that utilize embeddings to capture speaker characteristics and emotional states. For instance, VoiceLoop uses a phoneme-level language model to generate speech from text while preserving the speaker's voice by incorporating speaker-specific embeddings [3]. Similarly, Emotional TTS systems often rely on emotion embeddings to modulate the speech output to convey different emotional tones [4].

Contrastive Learning. The use of contrastive learning, as seen in Contrastive Language-Audio Pretraining (CLAP) [5], is a relatively new trend. Models like HuBERT and WavLM have shown that pretraining audio models on largescale unlabeled data using contrastive tasks can significantly improve the model's performance on downstream speech tasks [6; 7] by providing useful compressed latent representation which is easier to model than raw waveforms or spectrograms.

These related methodologies highlight the breadth of techniques employed in modern TTS systems, from end-to-end models to sophisticated generative networks using embeddings and contrastive learning. The convergence of these technologies represents a significant step forward in the quest for more natural and expressive synthetic speech.

Diffusion Models. Diffusion models have recently been explored as a powerful method for generating high-quality speech. These models, such as WaveGrad and DiffWave, use a gradual denoising process to synthesize speech, starting from noise and progressively refining the signal into intelligible speech [8; 9]. The process involves a learned reverse diffusion that transforms a Gaussian noise distribution into a complex signal [10]. Transformer-based Text-to-Speech Models. Transformer-based architectures have significantly influenced the development of TTS systems, offering substantial improvements over traditional methods. These models fall into two main categories: autoregressive and non-autoregressive models, each with unique attributes and applications in speech synthesis. Autoregressive Models: Autoregressive models, such as [11], generate signals sequentially, predicting one segment at a time based on all previously generated segments. This approach ensures high coherence and naturalness in the speech output. The transformer's attention mechanism allows these models to capture long-range dependencies in text, crucial for prosody and intonation in speech. A typical example includes the original Transformer TTS, which utilizes a self-attention mechanism to model temporal sequences in a highly parallelizable manner:

$$p(y|x;\theta_{AR}) = \prod_{t=0}^{T} p(y_t|y_{<t}, x;\theta_{AR}) \tag{2.1}$$

where y_t is the predicted audio output at time t, and x_t is the input phoneme or text sequence up to time t.

Non-autoregressive Models: In contrast, non-autoregressive models such as FastSpeech [12] bypass the sequential dependency of autoregressive models, predicting all parts of the speech output simultaneously. This leads to significantly faster synthesis times and reduces latency, which is beneficial for real-time applications. FastSpeech and its successors, like FastSpeech 2 [13], improve on this approach by predicting duration, pitch, and energy explicitly, which are then used to modulate the speech synthesis process.

$$\hat{y} = \text{Parallel Decoder}\big(\text{Duration Predictor}(x)\big) \tag{2.2}$$

where $\hat{y}$ represents the entire speech waveform generated in parallel, and x is the input phonetic/text representation.

Both types of transformer-based TTS models have pushed the boundaries of speech synthesis, offering more natural, flexible, and efficient solutions. However, the choice between autoregressive and non-

autoregressive approaches often depends on the specific requirements of latency, naturalness, output diversity and computational resources.

3. Model description

Masked Autoencoder for Feature Compression. The Masked Autoencoder (MAE) processes an input audio signal x by first computing its log mel spectrogram $X \in R^{T\times F}$, where T indicates the time steps, and F represents the mel frequency bins. This spectrogram X is analogized to an image and segmented into patches of size $P \times P$, where each patch size P is a divisor of both T and F. These patches are then input into the AudioMAE encoder. The encoder, a convolutional neural network, operates with a kernel and stride both set to P, producing an output with D channels. Consequently, the encoder output is $E \in R^{T'\times F'\times D}$, where $T' = \frac{T}{P}$ and, $F' = \frac{F}{P}$ and D is the dimension of the embedding produced by MAE. The encoded features E are treated as the latent representation for subsequent processing.

To train the AudioMAE, a loss function is employed, specifically the Mean Squared Error (MSE) loss, calculated over the masked patches to assess the reconstruction quality. The MSE loss is defined as:

$$\text{MSE Loss} = \frac{1}{N_{\text{masked}}}\sum_{i=1}^{N_{\text{masked}}}\left(\widehat{X_i} - X_i\right)^2 \tag{3.1}$$

where N_{masked} is the number of masked patches, X_i is the original patch, and $\widehat{X_i}$ is the reconstructed patch output by the decoder of the MAE.

Conditioning Information C: Reference Audio and Text Phonemes. In the audio generation model, the conditioning information C plays a crucial role in guiding the generative process by providing contextual cues that influence the output. For this model, C is derived from two primary sources: reference audio and text phonemes, each contributing unique aspects to the generation process.

CLAP Autoencoder for Conditioning. The CLAP autoencoder is designed to project both audio and text into a unified multimodal space, enabling the effective use of this information as conditioning data. Let Xa denote the processed audio, represented in a matrix $X_a \in R^{F\times T}$, where F is the number of spectral components, such as Mel bins, and T is the number of time bins. Similarly, let X_t denote the text representation. Within a batch of N audio-text pairs, these are denoted as $\{X_a, X_t\}$.

The audio and text data are encoded via separate encoder functions, $f_a(\cdot)$ and $f_t(\cdot)$ respectively. For a batch of N items, the encoded representations are given by:

$$\widehat{X_a} = f_a(X_a); \quad \widehat{X_t} = f_t(X_t) \tag{3.2}$$

where $\widehat{X_a} \in R^{N\times V}$ and $\widehat{X_t} \in R^{N\times U}$ represent the dimensionalities V and U of the audio and text representations, respectively.

To bring these representations into a joint multimodal space of dimension d, learnable linear projections are applied:

$$E_a = L_a\left(\widehat{X_a}\right) \tag{3.3}$$

$$E_t = L_t\left(\widehat{X_t}\right) \tag{3.4}$$

where $E_a, E_t \in R^{N\times d}$ are the projected embeddings for audio and text, and L_a, L_t are the respective linear projection functions.

The similarity between the audio and text embeddings is computed in the joint space as follows:

$$C = \tau \cdot (E_t \cdot E_a^{\top}) \tag{3.5}$$

where τ is a temperature parameter that scales the range of the logits. The similarity matrix $C \in R^{N\times N}$ includes correct pairs along the diagonal and incorrect pairs off the diagonal.

A symmetric cross-entropy loss is then computed over the similarity matrix to train the encoders and their projections:

$$\mathcal{L} = 0.5 \cdot \left(\mathrm{l}_{\text{text}}(C) + \mathrm{l}_{\text{audio}}(C)\right) \tag{3.6}$$

where $\mathrm{l} = \frac{1}{N}\sum_{i=0}^{N}\log\left(\text{softmax}(C)\right)$ along the text and audio axes, respectively. This loss function facilitates the joint training of the audio and text encoders, enhancing their capability to encode relevant features effectively for audio generation tasks.

Text phoneme encoding: Text phonemes represent another vital component of the conditioning information. Phonemes, the smallest units of sound in a language, are extracted from the input text and encoded to capture the linguistic nuances and articulatory features necessary for generating coherent and contextually appropriate audio. This encoding process transforms textual data into a sequence of phonetic representations, C_{phonemes}, which are then used to condition the audio generation, ensuring that the produced audio matches the intended linguistic content and style dictated by the input text.

Together, these conditioning components $C = \{C_{\text{ref}}, C_{\text{phonemes}}\}$ integrate multiple modalities—audio and text—providing a comprehensive set of cues that enhance the model's ability to generate high-fidelity and contextually rich audio outputs.

Autoregressive Modeling for Intermediate Representation. This model component is responsible for generating a latent representation from diverse conditioning information using an autoregressive approach inspired by transformer-based models. The formulation of the autoregressive model $\mathcal{M}_\theta$ is given by:

$$\hat{Y} = \mathcal{M}_\theta(C) \tag{3.7}$$

where C represents conditioning information, and $\hat{Y}$ is the predicted latent representation. The model $\mathcal{M}_\theta$, parameterized by θ, predicts the next sequence element based on previous ones, maximizing the probability distribution across the sequence:

$$\text{argmax}_\theta \prod_{i=1}^{L} P\left(y_i \middle| C_{\text{ref}}, C_{\text{phonemes}}, y_1, y_2, \dots, y_{i-1}; \theta\right) \tag{3.8}$$

where L is the length of the latent sequence Y encoded by MAE, and y_i are its components. Variational Autoencoder (VAE) for Diffusion Modeling: A Variational Autoencoder (VAE) [14] primarily for feature compression and to learn a compact audio representation, z, which is dimensionally much smaller than the original audio signal, x.

The operation of the VAE can be expressed through the forward pass equation:

$$\mathcal{V}: X \mapsto z \mapsto \hat{X} \tag{3.9}$$

where X represents the mel-spectrogram of the audio input x, and $\hat{X}$ is the reconstruction of X. This reconstructed spectrogram, $\hat{X}$, can subsequently be transformed back into the audio waveform $\hat{x}$ using a pretrained HiFiGAN vocoder [15].

To optimize the parameters of the VAE, a reconstruction loss and a discriminative loss are computed based on the comparison between X and $\hat{X}$. Furthermore, the VAE architecture employs a regularization strategy by computing the KullbackLeibler (KL) divergence between the latent representation z and a standard Gaussian distribution with mean $\mu = 0$ and variance $\sigma^2 = 1$:

$$\text{KL Loss} = D_{\text{KL}}\left(\mathcal{N}(z; \mu_z, \sigma_z^2) \parallel \mathcal{N}(0,1)\right) \tag{3.10}$$

This regularization helps to maintain the statistical properties of the latent space, ensuring that z adheres closely to a Gaussian distribution, thereby stabilizing the generation process and enhancing the quality of the reconstructed audio.

Latent diffusion model for audio synthesis. The audio synthesis is performed using a latent diffusion model that operates within the latent space provided by the VAE autoencoder. This model is expressed through a series of diffusion steps, starting with a latent representation z and gradually adding noise to reach a diffusion state z_T:

$$z_t = \sqrt{1 - \beta_t} z_{t-1} + \sqrt{\beta_t} \epsilon_t \tag{3.11}$$

where β_t is a noise schedule parameter, and $\epsilon_t \sim \mathcal{N}(0, I)$ is Gaussian noise.

The reverse process involves a gradual denoising of z_T to reconstruct the latent representation:

$$z_{t-1} = \frac{z_t - \sqrt{\beta_t}\epsilon_t}{\sqrt{1-\beta_t}} \tag{3.12}$$

The optimization targets the minimization of the difference between the original and reconstructed latent representations, defined by the loss function:

$$\mathcal{L}(\phi) = E_{z_{\mathbb{0}}, \epsilon \sim \mathbb{N}(\mathbb{0}, I), t}[|z_0 - \text{Dec}(z_t; \phi)|^2] \tag{3.13}$$

where ϕ are the parameters of the diffusion model, Dec denotes the decoding function of the diffusion model, and z_0 is the original latent representation.

Adaptation of Pretrained Model Components. In the development of our model, we utilized components from the pretrained Audio Latent Diffusion Model 2 (Audio LDM2). This approach allowed us to leverage the robust foundations established by the existing model, particularly its effective handling of complex audio data through diffusion processes. An important modification in our methodology involved the adaptation of the conditioning mechanism used in Audio LDM2. Traditionally, Audio LDM2 employs a conditioning vector C that incorporates CLAP-encoded text embeddings to guide the audio synthesis process. In contrast, our model replaces these text embeddings with CLAP-encoded audio embeddings which is intended to encode emotion and speaker information. This change aligns better with our focus on enhancing audio quality and relevance in text-to-speech applications, where the direct correlation between the input audio characteristics and the generated output is crucial.

$$C_{\text{ref}} = f_a(X_{ref}) \tag{3.14}$$

where C_{ref} represents the new conditioning vector using audio embeddings, $f_a(\cdot)$ is the CLAP audio encoder, and X_{ref} is the referebce audio feature matrix.

This adaptation not only tailors the model to our specific use case more closely, but also optimizes the interaction between the conditioning information and the generative components of the model. By integrating audio embeddings directly, our model gains a more nuanced understanding of the audio features, potentially leading to more accurate and lifelike audio generation in TTS systems.

4. Evaluation metrics

This section describes the evaluation process of our generative model.

The performance of the updated AudioLDM2 model is evaluated using several key metrics:

- Speaker Similarity: Quantifies the ability of the TTS system to preserve the unique characteristics of the speaker's voice.
- Emotion Classification Error: Measures the model's accuracy in conveying the intended emotional states in the synthesized speech.
- Word Error Rate (WER) / Character Error Rate (CER): Assesses the intelligibility and accuracy of the spoken output, comparing the transcribed text from the synthesized speech to the original input text.

All metrics reported below were calculated using the Amphion software [16] - a toolkit library for audio generation. These metrics provide a comprehensive framework for assessing the effectiveness of the TTS system in producing high-quality, emotionally expressive, and speaker-specific speech.

5. Results and metrics description

Speaker Similarity Metric. To quantitatively assess the speaker similarity between the reference and generated audio samples, we employed a speaker verification model based on WavLM, a state-of-the-art audio processing model [7]. This model was pretrained using a contrastive loss, which optimizes the embeddings to minimize the distance between similar pairs and maximize the distance for dissimilar pairs, making it well-suited for speaker verification tasks.

The metric we report is the average cosine similarity between the embeddings of reference audio samples and their corresponding generated samples. The embeddings are extracted using the WavLM model, which captures speaker-specific characteristics. Cosine similarity measures the cosine of the angle between two vectors in the embedding space, providing a scale from -1 (completely different) to 1 (identical), where higher values indicate greater speaker similarity. The formula for cosine similarity is given by:

$$\text{Cosine Similarity} = \frac{\sum_{i=1}^{n} A_i \cdot B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \cdot \sqrt{\sum_{i=1}^{n} B_i^2}} \tag{5.1}$$

where Ai and Bi are the components of the embeddings from the reference and generated audio samples, respectively. This metric effectively quantifies how well the generated audio preserves the identity characteristics of the speaker in the reference audio. The results are presented in Table 1

Table 1. Comparison of speaker similarity scores

Model	Speaker similarity
Proposed model	0.63
xTTS v2	0.9

Emotional Classification Accuracy. The accuracy of emotion recognition was evaluated using the Emotion2Vec model [17], which predicted emotions for both the reference and the produced audios. This measure reflects the model's ability to encode and reproduce the emotional states intended by the original speech. Results are tabulated in Table 2.

Table 2. Comparison of emotional classification accuracy

Model	Emotion classification accuracy
Proposed model	0.035
xTTS v2	0.17

Metrics for measuring WER and CER were calculated using transcripts generated by a pre-trained large Whisper [18] Automatic Speech Recognition (ASR) model, comparing these against the ground truth transcripts.

WER is computed as the ratio of the total number of operations (insertions, deletions, and substitutions) needed to convert the ASR-generated transcript into the ground truth transcript, divided by the total number of words in the ground truth transcript. The formula for WER is given by:

$$\text{WER} = \frac{S+D+I}{N} \tag{5.2}$$

where S is the number of substitutions, D is the number of deletions, I is the number of insertions, and N is the number of words in the ground truth transcript.

Similarly, CER is calculated by applying the same principle at the character level rather than the word level. It measures the minimum number of insertions, deletions, and substitutions required to change the ASR-generated transcript into the ground truth, normalized by the total number of characters in the ground truth transcript. The formula for CER is:

$$\text{CER} = \frac{s+d+i}{n} \tag{5.3}$$

where s represents substitutions, d represents deletions, i represents insertions, and n is the total number of characters in the ground truth transcript.

Both metrics provide crucial insights into the transcription accuracy of the generated speech, with lower values indicating higher accuracy and better performance of the text-to-speech synthesis system. The results are presented in Table 3.

Table 3. Comparison of Word error rate and Character error rate

Model	Word error rate↓	Character error rate↓
Proposed model	1.0	1.01
xTTS v2	0.21	0.02

5. Conclusions

This study provided the evaluation of the latent diffusion model, against the xTTS v2 model using a set of rigorous metrics on the EmoV-DB dataset. The findings revealed some insights into the performance of both models in terms of speaker similarity, emotional preservation, and intelligibility.

While the proposed model demonstrated decent efficiency in maintaining speaker characteristics, as indicated by the speaker similarity score, it was outperformed by xTTS v2 in all assessed metrics. Notably, our model exhibited considerable deficiencies in emotional classification accuracy, suggesting that the audio CLAP embeddings it relies on may be more attuned to capturing speaker-related information than the nuances of emotional expression. This observation was underscored by the model's

low emotion classification error rate, which points to a potential misalignment between the emotional intents encoded by the embeddings and those expressed in the speech output. Also, pretrained dataset for the CLAP component, which is a mix of speech and general audio and its corresponding captions, might not be effective for speech synthesis, suggesting that pre-training on transcribed speech dataset may improve generation quality.

REFERENCES

1. Y. Wang et al. Tacotron: Towards End-to-End Speech Synthesis. *Interspeech 2017. ISCA: ISCA*, 20-24 August 2017, Stockholm, Sweeden, 2017, p. 4006-4010
2. J. Shen et al. Natural TTS synthesis by conditioning wavenet on MEL spectrogram predictions. *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, 15–20 April 2018, Calgary, AB, Canada, 2018, p. 4779-4783
3. Taigman Y. Voiceloop: Voice fitting and synthesis via a phonological loop, 2018 (Preprint Arxiv:1707.06588)
4. Lee Y. Emotional end-to-end neural speech synthesizer, 2017 (Preprint Arxiv: 1711.05447)
5. Y. Wu, K. Chen, T. Zhang, Y. Hui, T. Berg-Kirkpatrick and S. Dubnov. Large-Scale contrastive language-audio pretraining with feature fusion and keyword-to-caption augmentation. *ICASSP 2023 - 2023 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, 4–10 June 2023, Rhodes Island, Greece. 2023
6. W.-N. Hsu et al. HuBERT: self-supervised speech representation learning by masked prediction of hidden units. *IEEE/ACM transactions on audio, speech, and language processing*. 2021. vol. 29. p. 3451–3460.
7. S. Chen et al. WavLM: large-scale self-supervised pre-training for full stack speech processing. *IEEE journal of selected topics in signal processing*. 2022. Vol. 16, p. 1505–1518.
8. Chen N. Wavegrad: Estimating gradients for waveform generation. *International Conference on Learning Representations (ICLR)*, 2020.
9. Kong Z. Diffwave: A versatile diffusion model for audio synthesis. *International Conference on Learning Representations (ICLR)*, 2021.
10. Chen M. An overview of diffusion models: Applications, guided generation, statistical rates and optimization, 2024 (Preprint Arxiv: 2404.07771)
11. Wang C. Neural codec language models are zero-shot text to speech synthesizers, 2023 (Preprint Arxiv: 2301.02111)
12. Ren Y. Fastspeech: Fast, robust and controllable text to speech. *Advances in Neural Information Processing Systems 32 (NeurIPS 2019)*, 8-14 December 2019, Vancouver Convention Centre, Canada, vol 32.
13. Ren Y. Fastspeech 2: Fast and high-quality end-to-end text to speech. *ICLR 2021 The Ninth International Conference on Learning Representations*, 2021
14. Kingma D. P. Auto-encoding variational bayes. *International Conference on Learning Representations*. 14-16 April 2014, Banff, AB, Canada, 2014
15. Kong J. et al. Hifi-gan: Generative adversarial networks for efficient and high-fidelity speech synthesis. *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*, 6-12 December, 2020, vol 33, p. 17022-17033.
16. Xueyao Zhang, Liumeng Xue, Yicheng Gu ett.al. Amphion: An open-source audio, music and speech generation toolkit, 2024 (Preprint Arxiv: 2312.09911)
17. Ma Z. et al. Emotion2vec: Self-supervised pre-training for speech emotion representation, 2023 (Preprint Arxiv: 2312.15185)
18. Radford A. Robust speech recognition via large-scale weak supervision. *Proceedings of Machine Learning Research*, 23-29 July, 2023, vol 202, p. 28492-28518.

Іванюк Андрій Олегович *Аспірант докторської школи*
Національний університет "Києво-Могилянська академія", факультет інформатики, вулиця Григорія Сковороди 2, Київ, Україна, 04655
e-mail: a.ivaniuk@ukma.edu.ua
https://orcid.org/0000-0002-4189-3787

Модель латентної дифузії для обробки мовного сигналу

Актуальність. Розробка генеративних моделей для синтезу аудіо, включаючи текст-у-мовлення (англ. text-to-speech, TTS), текст-у-музику та текст-у-аудіо застосування, значною мірою залежить від їх здатності обробляти складні та різноманітні вхідні дані. В цій роботі ми розглядаємо латентне дифузійне моделювання - універсальний підхід, який використовує стохастичні процеси для генерації високоякісних аудіо сигналів.

Мета. Це дослідження має на меті оцінити ефективність латентного дифузійного моделювання для аудіо синтезу на основі набору даних EmoV-DB, який містить записи з багатьма мовцями, з п'ятьма емоційними станами, та порівняти його з іншим генеративними методом.

Методи дослідження. Ми застосували латентне дифузійне моделювання спеціально для синтезу мовлення та оцінили його ефективність за допомогою метрик, які визначають зрозумілість, подібність голосу та збереження емоцій в згенерованому аудіо сигналі.

Результати. Дослідження показує, що запропонована модель демонструє пристойну ефективність у збереженні характеристик голосу, але поступається дискретній авторегресивній моделі: xTTS v2 за всіма оціненими метриками. Зокрема, досліджувана модель виявляє недоліки в точності класифікації емоцій, що вказує на можливе невідповідність між емоційними намірами, закодованими у векторах, та тими, що виражені у згенерованому сигналі.

Висновки. Результати вказують на те, що подальше вдосконалення здатності нейронної мережі кодувальника обробляти та інтегрувати емоційні дані покращує ефективність латентної дифузійної моделі. В наших подальших дослідженнях ми плануємо зосередитися на оптимізації балансу між характеристиками мовця та емоційними характеристиками в TTS моделях для досягнення більш цілісного та ефективного синтезу людського мовлення.

Ключові слова: аудіо моделювання, штучні нейронні мережі, синтез мовлення.

Надійшла у першій редакції 02.03.2024, в останній - 11.04.2024.
The first version has been received on 02.03.2024, the final version - on 11.04.2024.

DOI: https://doi.org/10.26565/2304-6201-2024-61-06

УДК (UDC) 519.7, 53.02, 007.5

Obraztsov Dmytro Ihorovych *student*
V.N. Karazin Kharkiv National University, 4 Svobody Square, Kharkiv, Ukraine, 61022.
e-mail: xa12850498@student.karazin.ua
https://orcid.org/0009-0002-3479-1093

Yanovsky Volodymyr Volodymyrovych *Doctor of Physical and Mathematical Sciences, professor*
V. N. Karazin Kharkiv National University, 4 Svobody Square, Kharkiv, Ukraine, 61022;
Institute of Single Crystals, National Academy of Sciences of Ukraine, 60 Nauky Ave., Kharkiv, Ukraine, 61001.
e-mail: yanovsky@isc.kharkov.ua;
https://orcid.org/0000-0003-0461-749X

The appearance of «intelligence» in self-propelled bots

Relevance. Nowadays, the study of the behavior and properties of active matter, which corresponds to the collective behavior of self-moving elements, is very promising field of research. Active matter is widespread in nature and is used in various modern technologies.
Objective. To study the collective behavior of mobile bots in a simple maze and to determine the distinctive trends in the single bots exiting from the maze as their number increases.
Research methods. To perform the research, mobile bots have been created and their behavior in the maze has been analyzed. Their positions and interactions were recorded on a video, and processed to obtain the necessary data.
Results. The research revealed the existence of an optimal number of bots in the maze for which the average time to exit the maze is minimal. The determined dependence of the probability of bots leaving the maze is non-monotonic, and there is a number of bots for which this probability is minimal. It has been determined that with 12 bots in the maze, both the average exit time and the exit probability are minimal. Thus, with this number of bots, a small number of bots quickly exit the maze. A quantitative measure of bot intelligence, the intelligence coefficient, is proposed. There exists an optimal number of bots that maximizes the measure of «intelligence» concerning the task of exiting from the maze. Both a decrease and an increase in the number of bots lead to a reduction in the «intelligence» of the bot collective. The measure of the «intelligence» of a bot collective surpasses that of an individual bot.
Conclusions. In this work, we have considered the groups of self-moving bots exiting the maze and the typical quantitative characteristics that allowed us to determine the main dependencies of their behavior.

***Keywords**: active matter systems, clustering, research, observation, visualization of results, arenas, bots, research application.*

How to quote: D.I. Obraztsov, and V.V. Yanovsky, "The appearance of «intelligence» in self-propelled bots." *Bulletin of V.N. Karazin Kharkiv National University, series "Mathematical modelling. Information technology. Automated control systems"*, vol. 61, pp. 52-60, 2024. https://doi.org/10.26565/2304-6201-2022-54-01

Як цитувати: D.I. Obraztsov, and V.V. Yanovsky, "The appearance of «intelligence» in self-propelled bots. *Вісник Харківського національного університету імені В.Н.Каразіна, серія Математичне моделювання. Інформаційні технології. Автоматизовані системи управління*. 2024. вип. 61. С.52-60. https://doi.org/10.26565/2304-6201-2022-54-01

1. Introduction

Recently, there has been considerable interest in the study of active matter. Active matter is understood as sets of self-moving elements or bots [1]. These systems exhibit special behavior that distinguishes them from ordinary systems of physical particles. The study of active matter began with the study of swarming behavior in fish [2]. In a more recent and significant work [3], the motion of bird-like objects, bots, has been modeled. Over time, both experimental and theoretical studies have deepened the understanding of behavioral mechanisms of various biological systems, as discussed in the review [4]. Obviously, artificial swarms of self-propelled bots, which are also actively investigated (see, e.g., [5]), belong to the category of active matter. In addition, such studies contribute to the field of cybernetics. An

example is the study of the interaction of abstract automata or other computational models. These studies are closely related to biological studies of animal behavior. The main goal of these studies has been to understand the behavior of different animals and their problem-solving abilities, which is akin to studying their "intellectual" abilities. In a sense, this direction is equally relevant to the creation of artificial systems with purposeful behavior.

The interaction of abstract automata, without loss of generality, can be regarded as the problem of guiding a finite automaton through a maze. One of the earliest works in this area was Shannon's study [6], which explored a maze-solving problem using a mouse automaton. This idea for this research emerged from existing and highly advanced biological studies of the behavior of various animals and creatures in mazes, with mice being classical subjects.

Studies on the behavior of automata in mazes have been further developed in works [7, 8], where questions about the existence of a finite automaton capable of traversing all possible mazes were formulated. It was quickly proven that such finite automaton does not exists [9, 10]. Further research has focused on determining which mazes could be traversed by a single finite automaton [7, 8, 11, 12], and some generalizations of automata, such as those with external memory, have also been considered (see, e.g., [13]).

In this work, we examine the behavior of self-propelled bots in a simple labyrinth, inspired by the framework proposed in [14] where the exit of conventional elements from the maze has been studied in detail. For our research, we constructed self-propelled bots with micro-vibration motors. Subsequent experiments to navigate the bots through the maze allowed us to identify key dependencies that characterize the behavior of bots when exiting the maze. We introduced the concept of "intelligence coefficient" as the ratio of the average exit time of a single bot to the average exit time of all bots from the maze. Our results indicate the existence of a certain number of bots that optimally solve the problem of exiting the maze in minimum time.

2. Self-propelled bots

For the experiments, a self-propelled bot with a vibration motor has been developed (Fig. 2.1). Unlike the previously used variants [15, 16], where the motion of the bots was stimulated by vibrations of the external surface, each bot was equipped with a vibration motor and power source. Such a design required multi-parameter optimization of the power source parameters and placement, as well as the optimization of the vibration motor, to achieve stable mobility of the bots. The advantages of this design include the autonomy of each bot and its relatively high independence of.

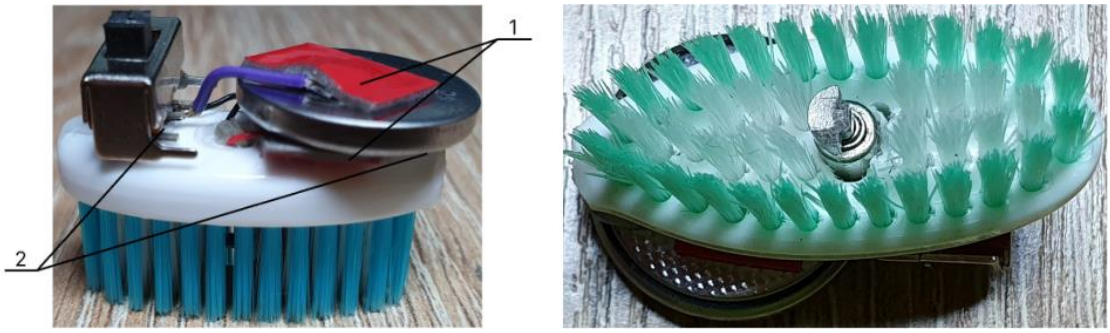


Fig. 2.1 On the left is the overall view of the bot: 1 – updated adhesive strip; 2 – adhesive layer for fixation, switch, and wire connections. On the right the motor placement is presented

A microvibration motor with the following characteristics has been chosen to propel the bot:

- Motor length (total): 8 (11) mm;
- Wire length: 10 mm;
- Weight: 0.6 g;
- Voltage: 1.5 V, Current: 15 mA;
- Voltage: 3 V, Current: 23 mA.

The bot's power source is a lithium battery with the following characteristics:

- Model: CR2032;
- Voltage: 3 V;
- Battery capacity, mAh: 220;
- Weight: 2.9 g.

Due to the shifting center of gravity, the elliptical platform with bristles can move relatively rectilinearly on a flat surface. After considering several options, a prototype in which the motor was located inside the body of the bot allowing maximum transmission of vibration impulses to the locomotive part of the bot has been selected (Fig. 2.1). In order not to interfere with the rotational motion of the motor, the bristles around the motor location have been removed.

Fig. 2.1 shows the location of the motor inside the bot body. The isolated bristle at the motor location can be seen, and it is clear that it does not interfere with the rotational motion of the microvibration motor. The characteristic dimensions of the bot are $1{,}8 \times 3{,}5\ cm^2$ with a height of 2.5 cm.

These last improvements completed the bot design. After that, replication of the bots began so that the total number of bots would be at least twenty-one units (fifteen main bots and a minimum of five spare bots in case of malfunction). Additional bots were necessary because for the proposed modification, it is easier to replace a bot than, to replace a battery, for example.

After launching the bots, a characteristic interaction characterized as elastic collision was observed during each experiment. Upon collision, the bots formed clusters, demonstrating signs of system self-organization. However, as the study showed, the bot clusters were quite dynamic and constantly changed their positions in the constructed arenas, as well as the number of agents (bots) in them.

3. Experimental observations

After creating the necessary number of bots, a simple maze has been constructed (Fig. 3.1). The outer boundary of the maze has been made of wood, and the internal intersections from foam. The glass surface has been used for the bots to move on. The dimensions of the maze that need to be characterized are the area of the wooden boundary, the structural elements of the maze, and the size of the maze exit. Therefore, the area of the maze outline is $S_{outline} = 1046.3467\ cm^2$; the area of the T-shaped and rectangular elements is $S_{T-element} = 77.4\ cm^2$ and $S_{rectangle} = 19.6\ cm^2$, respectively. According to the defined dimensions, the width of the maze exit was determined to be 9.5 cm.

In each experiment, an initial set of bots was placed in a relatively small area measuring $8.2 \times 11.5\ cm^2$, and the initial movement directions were random. It is worth noting that the bots started from containers located in the middle of the maze (Fig. 3.1).

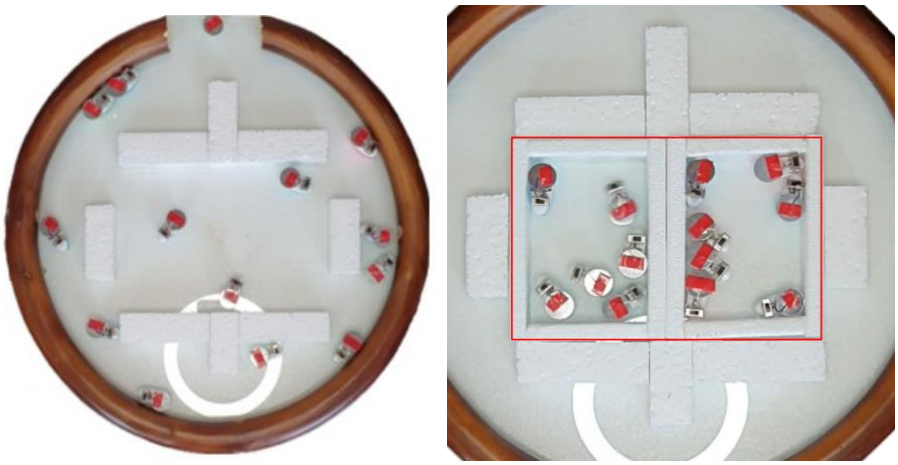


Fig. 3.1 On the left is a general view of the maze arena. On the right is a demonstration of containers for bots

The behavior of the bots has been recorded on video. Quantitative and qualitative details of the bot's behavior in the maze were determined from the video recordings. Three observations were conducted for each set of bots (2 – 15).

4. Experimental results

Observations over groups of bots allowed us to obtain data on the dynamics of bots exiting the maze.

Starting from one bot, the average time of bots exiting the maze was determined. This time was calculated in accordance with the relation (4.1):

$$\langle t\rangle_N = \frac{1}{n}\sum_{i=1}^{n} t_i \tag{4.1}$$

where t_i is the exit time of the i-th bot from the labyrinth;

N is the initial number of bots in the labyrinth;

$n < N$ is the number of bots from N that found the labyrinth exit.

Experimental data for t_i have been determined by timing video recordings. It is important to note that averaging was performed only for bots that exited the labyrinth. Otherwise, the exit time would depend on the waiting time, which is not acceptable. After conducting several experiments (three in particular), these average values were additionally averaged over realizations. Mean square errors for these data were also determined. The results for the average exit time of bots depending on the number of bots N are shown in Fig. 4.1.

In Fig. 4.1, the dots represent experimental data, connected by straight lines for clarity. Mean square errors are indicated by vertical bars.

Determining the average exit time for a single bot posed a certain problem. The reason is that a solitary bot does not always find the exit from the labyrinth. To determine $\langle t\rangle_1$, additional experiments were conducted (20 observations), in which the exit time of a single bot from the labyrinth was recorded when there were no other bots in the labyrinth. Only three observations from the ensemble were used because it took 20 experiments to record the bot's exit within 40 seconds for three of them. Thus, the probability of a solitary bot exiting the labyrinth is quite small $p_1 = 3/20$. It should be noted that for a solitary bot, exiting the labyrinth is a challenging task.

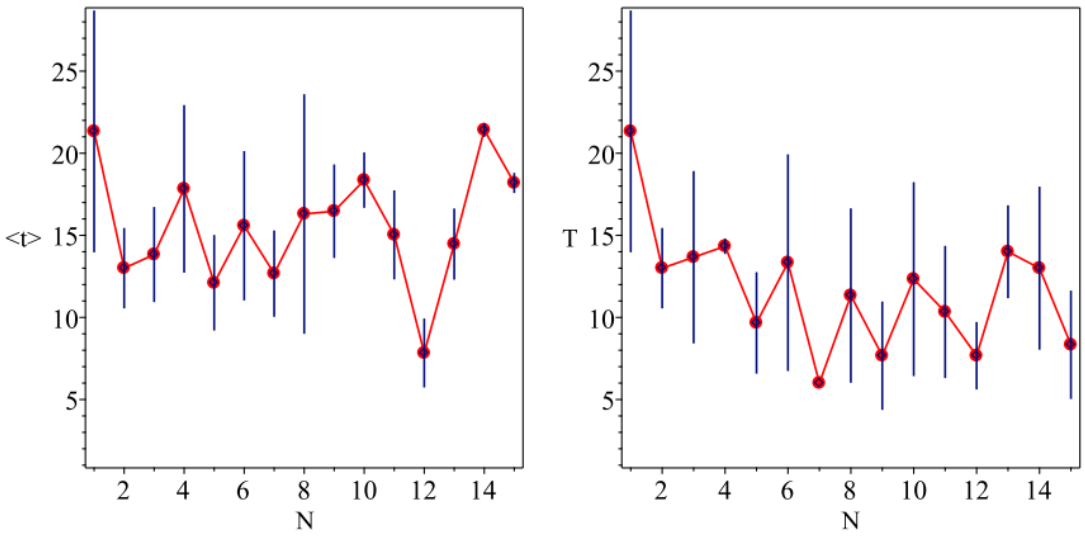


Fig. 4.1 On the left, the dependence of the average exit time of a bot from the labyrinth on the number of bots in it is presented. On the right, the average exit time of the first bot from the labyrinth and the data errors are shown as well, depending on the group size

It is easy to notice the presence of a certain number of bots (12), for which the average exit time from the labyrinth is minimal.

It is interesting to consider the average exit time of the first bot as an additional characteristic. This dependence on the number of bots in the group is shown in Fig. 4.1 on the right. It is easy to notice that this dependence does not have a pronounced minimum. The minimum value is achieved with a different number of bots (7) in the group. Errors for these data also noticeably increase. Thus, the exit time of the first bot weakly depended on the number of bots in the group. The solitary bot is an exception, as its exit time is significantly longer. Therefore, the average exit time of all bots from the labyrinth should be used as a value sensitive to the number of bots in the group.

Another important characteristic of the behavior of groups of bots in the maze is the probability of

bots leaving the maze. We define it as the ratio of the number of bots that left the maze to the total number of bots. Based on the data obtained, we determined the probability of a bot exiting the maze as a function of the initial number of bots in the maze. This dependence presented on the Fig. 4.2.

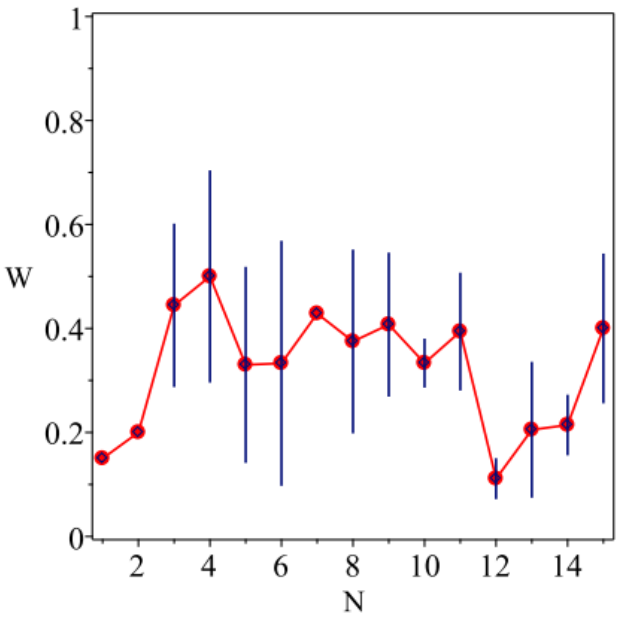


Fig. 4.2 Probability of a bot belonging to a group of N bots to exit the labyrinth

It is easy to see that a solitary bot has a low probability of exiting. It should be noted that its probability was determined differently for understandable reasons. The probability was taken as the relative frequency of favorable events, i.e., the number of experiments in which the solitary bot exited the labyrinth, to the total number of experiments. From Fig. 4.2, it can be deduced that a group of bots more efficiently solves the task of exiting the labyrinth than a solitary bot. It is important to note that there is a probability minimum when the number of bots is 12, the same number for which the average time for bots to exit also has a minimum. In fact, this means that bots exit quickly but with low probability, or a relatively small fraction of the group. In a sense, this can be interpreted as a kind of "altruism" - the group ensures that a small part of the group exits the maze quickly.

Let's return to the discussion of the overall task of finding labyrinth exit. In biology, and beyond, this task has often been employed to assess the ability of a subject to solve the maze exit problem. Comparing the performance of different entities in this task allows conclusions to be drawn about their abilities to address this challenge, essentially making insights about their «intellectual» capabilities. In this context, «intellect» is generally understood as the «ability to solve problems». While there is no universally accepted definition of intelligence, most definitions focus on human intelligence [17]. However, for the broadest working definition, we will rely on the one provided above. While this definition has obvious limitations, it serves well as a characterization of the abilities of animals, humans, and swarm intelligence as well.

Based on this, it is natural to introduce a quantitative measure to determine intelligence based on how the labyrinth exit task is solved. If a collective solves the exit task in less time than another, it is deemed more «intellectual». It is logical to measure intelligence in a swarm relative to its solitary representative. Therefore, we will use the ratio of the average exit time for a solitary bot to the average exit time for the swarm as a coefficient of «intelligence» (Fig. 4.2):

$$I_N = \frac{\langle t \rangle_1}{\langle t \rangle_N} \tag{4.2}$$

where $\langle t \rangle_1$ – average exit time for an individual bot from the labyrinth;

$\langle t \rangle_N$ – average exit time for a system of N bots.

It could be seen that, if the average exit time for a swarm is less than the exit time for an individual bot, then the intelligence coefficient will exceed one. For a solitary bot, its value equals one by definition.

The dependence of the «intelligence» coefficient on the number of bots in the swarm is derived from the experimental data and presented on Fig. 4.3. It is noticeable that the intelligence coefficient for swarms exceeds its value for a solitary bot within the margin of error. It is interesting to note the presence of a clear maximum value $I_{12} \approx 2.7$ when there are 12 bots in the swarm. Understandably, such a dependency corresponds to the behavior of bots in this particular labyrinth. Expectations are that changing the labyrinth may alter this relationship. In a sense, this implies a change in the task, and consequently, the intellectual abilities of the swarm may differ. However, general properties, such as the presence of a maximum, are likely to remain, albeit in a different location. This could be interpreted as the existence of an optimal number of bots in a swarm that solves a specific task most efficiently. Increasing the number of bots sharply decreases the intelligence coefficient. One possible reason for this is the formation of clusters or grouped bots that block the exit from the labyrinth.

Understandably, expanding the scope of tasks, i.e., including any labyrinths from which bots can exit, should cause this dependence to lose the presence of a maximum. This is inferred from the existence of a maximum for a specific labyrinth at a certain number of bots in the swarm. Accordingly, for several labyrinths, one should expect the presence of multiple maxima. Averaging over implementations will reduce their amplitude. As the number of labyrinths increases, there will always be one that can be successfully navigated with a larger number of bots in the swarm. It is possible to hypothesize that the intelligence coefficient concerning the traversal of any labyrinth will increase with an increase in the number of bots in the swarm.

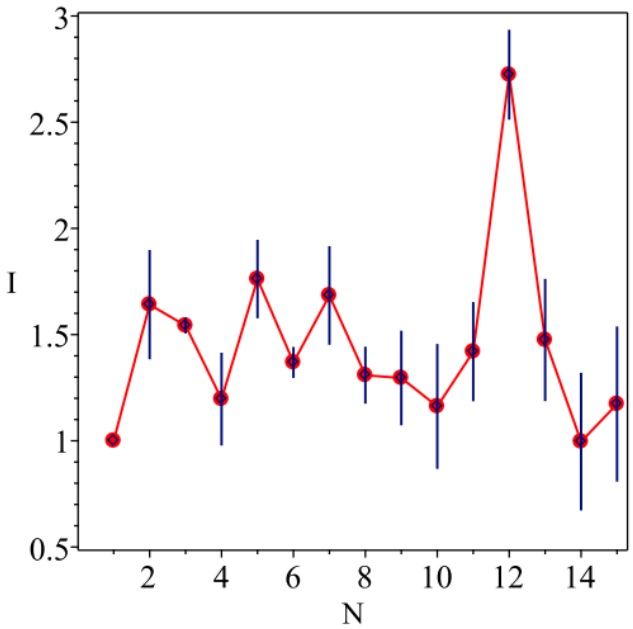


Fig. 4.3 Dependence of the intelligence coefficient on the number of bots in the swarm

5. Conclusions

In this study, the behavior of elementary bots in a labyrinth has been explored. Each self-propelled bot lacks specialized information processing systems and interacts solely through collisions with other bots, potentially leading to a statistical alignment of their behavior, a phenomenon that warrants further investigation. The research revealed the influence of the number of bots on the average exit time from the labyrinth. A distinctive dependence of the average exit time on the number of bots was identified (Fig. 4.1). The existence of an optimal number of bots in the labyrinth, resulting in the minimum average exit time, was determined. A non-monotonic dependence of the probability of bots exiting the labyrinth was obtained, with a specific number of bots minimizing this probability (Fig. 4.2). It was found that with 12 bots in the labyrinth, both the average exit time and the exit probability are minimal, signifying a quicker exit for a smaller portion of bots.

A quantitative measure of the «intelligence» of bots in solving the task of labyrinth exit was proposed. Naturally, the intelligence coefficient utilizes the ratio of the average exit time for a solitary bot to the average exit time for the swarm of bots. An optimal number of bots was identified, maximizing the

«intelligence» concerning the task of labyrinth exit performance (Fig. 4.3). Both a decrease and an increase in the number of bots result in a decrease in the swarm's «intelligence». The intelligence measure of the swarm surpasses that of the solitary bot, accounting for data errors.

It is essential to note that, at first glance, characterizing the efficiency of solving the labyrinth exit task by the fraction of bots that successfully exit may seem plausible. However, such a definition has significant drawbacks. For instance, if a labyrinth is chosen where a solitary bot always finds the exit, its exit success rate would be 1, a priori higher than the rate for other swarms. Consequently, such a characteristic does not align with the existence of swarm intelligence, observed and thus cannot be deemed acceptable.

СПИСОК ЛІТЕРАТУРИ

1. S. Ramaswamy, «The Mechanics and Statistics of Active Matter». Annual Review of Condensed Matter Physics, 2010. p.323–345. URL: https://www.annualreviews.org/doi/abs/10.1146/annurev-conmatphys-070909-104101
2. Aoki A simulation study on the schooling mechanism in fish. Bulletin of the Japanese Society of Scientific Fisheries, 48(8), 1982. p.1081-1088. URL: https://www.jstage.jst.go.jp/article/suisan1932/48/8/48_8_1081/_article
3. W. Reynolds. Flocks, herds, and schools, A distributed behavioral model, In Computer Graphics, pages 25-34, 1987. URL: https://dl.acm.org/doi/10.1145/37402.37406
4. T.Vicsek, A.Zafeiris, Collective motion. Physics Reports. 517 (3), 2012. p.71–140. URL: https://arxiv.org/abs/1010.5017
5. M.Rubenstein, A.Cornejo, R.Nagpal, Robotics. Programmable self-assembly in a thousand-robot swarm, Science, 345(6198), p.795-9, 2014. URL: https://pubmed.ncbi.nlm.nih.gov/25124435/
6. Shannon C.E., Presentation of a maze-solving machine, Cybernetics. Trans of the 8th conf. of the Josian Macy Jr. Found (Ed-H. v. Foerster), Nl, 1952. p.173-180. URL: https://www.kuenzigbooks.com/pages/books/28624/claude-shannon-elwood/presentation-of-a-maze-solving-machine-reproduced-paper
7. K.Dopp, Automaten in labirinthen I, Electronische Informationsverarbeitung und Kybernetik, v.7, No.2, 1971. p.79-94. URL: https://dblp.org/rec/journals/eik/Dopp71
8. K.Dopp, Automaten in labirinthen II, Electronische Informationsverarbeitung und Kybernetik, v.7, No.3, 1971. p.167-190. URL: https://dblp.org/rec/journals/eik/Dopp71a
9. H.Muller, Automata catching labyrinths with at most three components, Elektronische Informationsverarbeitung und Kybernetik, v.15, No.1/2, 1979. p.3-9. URL: https://dblp.org/rec/journals/eik/Muller79
10. H.Antelmann, L.Budach, H.A.Rollik, On universale traps, Elektronische Informationsverarbeitung und Kybernetik, v.15, No.3, 1979. p.123-131. URL: https://dblp.org/rec/journals/eik/AntelmannBR79
11. G.Asser, Bemerkungen zum labirinth-problem, Elektronische Informationsverarbeitung und Kybernetik, v.13, No.4,5, 1977. p.203-216. URL: https://dblp.org/rec/journals/eik/Asser77
12. R.Danecki, M.Karpinski, Decidability results on plane automata searching mazes, Proc. 2nd Int. FCT'7-9 Berlin: Conf. Akademie Verlag, 1979. p.84-91. URL: https://dblp.org/rec/conf/fct/DaneckiK79
13. M.Blum , C.Hewitt, Automata on a 2-dimensional tape, IEEE Conference Record, 8th Annual Symposium on Switching and Automata Theory, 1967. p.155-160. URL: https://doi.org/10.1109/FOCS.1967.6
14. D.M. Naplekov, V. V. Yanovsky, Thin structure of transit time distributions of open billiards, Phys. Rev. E 97, 012213(8), 2018. URL: https://doi.org/10.1103/PhysRevE.97.012213
15. D. Yamada, T. Hondou, and M. Sano, Coherent dynamics of an asymmetric particle in a vertically vibrating bed, Phys. Rev. E 67, 040301R (2003). URL: https://doi.org/10.1103/PhysRevE.67.040301
16. J.Deseigne, O.Dauchot, R.Chaté, Collective Motion of Vibrated Polar Disks, Physical Review Letters. 105 (9), 2010. URL: https://doi.org/10.1103/PhysRevLett.105.098001
17. Rita L. Atkinson, Richard C. Atkinson, Edward E. Smith, Daryl J. Bem, Susan Nolen-Hoeksema.

"Hilgard's Introduction to Psychology. History, Theory, Research, and Applications", 13th ed., 2000. URL: https://invent.ilmkidunya.com/images/Section/introduction-to-psychology-css-psychology-book.pdf

REFERENCES

1. S.Ramaswamy, «The Mechanics and Statistics of Active Matter». Annual Review of Condensed Matter Physics, 2010. p.323–345. URL: https://www.annualreviews.org/doi/abs/10.1146/annurev-conmatphys-070909-104101
2. Aoki A simulation study on the schooling mechanism in fish. Bulletin of the Japanese Society of Scientific Fisheries, 48(8), 1982. p.1081-1088. URL: https://www.jstage.jst.go.jp/article/suisan1932/48/8/48_8_1081/_article
3. W. Reynolds. Flocks, herds, and schools, A distributed behavioral model, In Computer Graphics, pages 25-34, 1987. URL: https://dl.acm.org/doi/10.1145/37402.37406
4. T.Vicsek, A.Zafeiris, Collective motion. Physics Reports. 517 (3), 2012. p.71–140. URL: https://arxiv.org/abs/1010.5017
5. M.Rubenstein, A.Cornejo, R.Nagpal, Robotics. Programmable self-assembly in a thousand-robot swarm, Science, 345(6198), p.795-9, 2014. URL: https://pubmed.ncbi.nlm.nih.gov/25124435/
6. Shannon C.E., Presentation of a maze-solving machine, Cybernetics. Trans of the 8th conf. of the Josian Macy Jr. Found (Ed-H. v. Foerster), Nl, 1952. p.173-180. URL: https://www.kuenzigbooks.com/pages/books/28624/claude-shannon-elwood/presentation-of-a-maze-solving-machine-reproduced-paper
7. K.Dopp, Automaten in labirinthen I, Electronische Informationsverarbeitung und Kybernetik, v.7, No.2, 1971. p.79-94. URL: https://dblp.org/rec/journals/eik/Dopp71
8. K.Dopp, Automaten in labirinthen II, Electronische Informationsverarbeitung und Kybernetik, v.7, No.3, 1971. p.167-190. URL: https://dblp.org/rec/journals/eik/Dopp71a
9. H.Muller, Automata catching labyrinths with at most three components, Elektronische Informationsverarbeitung und Kybernetik, v.15, No.1/2, 1979. p.3-9. URL: https://dblp.org/rec/journals/eik/Muller79
10. H.Antelmann, L.Budach, H.A.Rollik, On universale traps, Elektronische Informationsverarbeitung und Kybernetik, v.15, No.3, 1979. p.123-131. URL: https://dblp.org/rec/journals/eik/AntelmannBR79
11. G.Asser, Bemerkungen zum labirinth-problem, Elektronische Informationsverarbeitung und Kybernetik, v.13, No.4,5, 1977. p.203-216. URL: https://dblp.org/rec/journals/eik/Asser77
12. R.Danecki, M.Karpinski, Decidability results on plane automata searching mazes, Proc. 2nd Int. FCT'7-9 Berlin: Conf. Akademie Verlag, 1979. p.84-91. URL: https://dblp.org/rec/conf/fct/DaneckiK79
13. M.Blum , C.Hewitt, Automata on a 2-dimensional tape, IEEE Conference Record, 8th Annual Symposium on Switching and Automata Theory, 1967. p.155-160. URL: https://doi.org/10.1109/FOCS.1967.6
14. D.M. Naplekov, V. V. Yanovsky, Thin structure of transit time distributions of open billiards, Phys. Rev. E 97, 012213(8), 2018. URL: https://doi.org/10.1103/PhysRevE.97.012213
15. D. Yamada, T. Hondou, and M. Sano, Coherent dynamics of an asymmetric particle in a vertically vibrating bed, Phys. Rev. E 67, 040301R (2003). URL: https://doi.org/10.1103/PhysRevE.67.040301
16. J.Deseigne, O.Dauchot, R.Chaté, Collective Motion of Vibrated Polar Disks, Physical Review Letters. 105 (9), 2010. URL: https://doi.org/10.1103/PhysRevLett.105.098001
17. Rita L. Atkinson, Richard C. Atkinson, Edward E. Smith, Daryl J. Bem, Susan Nolen-Hoeksema. "Hilgard's Introduction to Psychology. History, Theory, Research, and Applications", 13th ed., 2000. URL: https://invent.ilmkidunya.com/images/Section/introduction-to-psychology-css-psychology-book.pdf

Образцов Дмитро Ігорович
студент магістратури
Харківський національний університет ім. В.Н. Каразіна, майдан Свободи, 4, Харків, Україна , 61022.
e-mail: xa12850498@student.karazin.ua
https://orcid.org/0009-0002-3479-1093

Яновський Володимир Володимирович
д. ф-м. н., професор
Харківський національний університет ім. В.Н. Каразіна, майдан Свободи, 6, Харків, Україна , 61022.
Інститут монокристалів, Національна Академія Наук України, проспект Науки, 60., Харків, Україна, 61001.
e-mail: yanovsky@isc.kharkov.ua
https://orcid.org/0000-0003-0461-749X

Виникнення «інтелекту» у саморухомих ботів

Актуальність. Наразі є дуже перспективним розвиток сучасного напрямку досліджень поведінки та властивостей активної матерії, яка відповідає колективній поведінці саморухомих складових. Така активна матерія широко поширена в природі та використовується у різних сучасних технологіях.
Мета. Провести дослідження колективної поведінки рухомих ботів у простому лабіринті та визначити характерні закономірності у виході цих частинок з лабіринту зі збільшення їх кількості.
Методи дослідження. Для виконання досліджень були створені рухомі боти та проведені експерименти по їх поведінці в лабіринті. Фіксація їх положень та взаємодії фіксувались відео записом, обробка якого дозволила отримати необхідні дані.
Результати. Виявлено існування оптимальної кількості ботів у лабіринті, для яких середній час виходу з лабіринту мінімальний. Визначена залежність ймовірності виходу ботів з лабіринту теж має не монотонний характер та існує кількість ботів за якої ця імовірність мінімальна. Визначено, що при наявних дванадцятьох ботах у лабіринті середній час виходу мінімальний, як і ймовірність виходу. Таким чином, при цій кількості, з лабіринту швидко виходить менша частина ботів. Запропоновано кількісну міру інтелектуальності ботів — коефіцієнт інтелекту. Існує оптимальна кількість ботів, яка має максимальну міру «інтелекту» по відношенню до задачі виходу ботів з лабіринту. Зменшення, як і збільшення кількості ботів веде до зменшення «інтелекту» колективу ботів. Міра «інтелекту» колективу ботів перевищує «інтелект» одного бота.
Висновки. В роботі було розглянуто вихід зграй саморухомих ботів з лабіринту та характерні числові характеристики, які дозволили визначити головні залежності їх поведінки.

***Ключові слова**: системи активної матерії, кластеризація, дослідження, спостереження, візуалізація результатів, арени, боти, застосунок для дослідження.*

Надійшла у першій редакції 07.03.2024, в останній - 13.04.2024.
The first version has been received on 07.03.2024, the final version - on 13.04.2024.

DOI: https://doi.org/10.26565/2304-6201-2024-61-07

УДК 004.8

Подоляка Оксана Олександрівна
к. т. н., доцент
Харківський національний університет імені В. Н. Каразіна,
площа Свободи 4, Харків, Україна, 61022
e-mail: podoliaka@karazin.ua;
https://orcid.org/0000-0002-3401-2996

Подоляка Олексій Миколайович
старший викладач
Харківський національний університет імені В. Н. Каразіна,
площа Свободи 4, Харків, Україна, 61022
e-mail: alex.podolyaka@gmail.com;
https://orcid.org/0000-0002-5755-3728

Оцінка корисності публічного набору даних для аналітичних досліджень

Організації та агенції публікують різні дані, які призначені для аналізу, навчання систем штучного інтелекту та інших дослідницьких цілей. Відповідно до прийнятих регуляцій у сфері захисту персональних даних публічні дані мають бути знеособлені та захищені від різних загроз розкриття персональних даних. Усунення цих загроз реалізується шляхом зменшення точності даних під час підготовки публічних даних. Втрата точності, вочевидь, призводить до зменшення корисності даних для аналізу. У роботі розглядаються ентропійні метрики корисності та проблеми їх обчислюваності, а також метрики втрати корисності окремих підмножин публічних даних.

Мета. Розробка ефективних метрик оцінки корисності публічного набору даних для аналізу з урахуванням вимог захисту персональних даних.

Методи дослідження. Інформаційна безпека, теорія інформації Шенона, управління даними (Data Governance).

Результати. Запропоновані метрики оцінки втрат інформації та корисності даних для аналізу на основі ентропійних метрик теорії інформації Шеннона. Запропоновано процедури, спрямовані на підвищення швидкодії обчислень розглянутих метрик.

Висновки. Описано процедури побудови безпечного публічного набору даних. Розглянуто питання застосування ентропійних метрик теорії інформації Шеннона для оцінки втрат інформації та корисності даних для аналізу. Показано, що обчислення зазначених метрик є складною, практично не здійсненною для великих баз даних, обчислювальною задачею. Запропоновано процедури, спрямовані на підвищення швидкодії обчислень розглянутих метрик. А саме, створення менш точної копії вихідних даних та формування випадкової вибірки із великої бази даних для обчислення необхідних статистик. Розглянуто метрики оцінки корисності для окремих підмножин (кластерів) публічних даних.

Ключові слова: *конфіденційність, деідентифікація, публікація даних, корисність даних, GDPR (General Data Protection Regulation).*

Як цитувати: Подоляка О. О, Подоляка О. М. Оцінка корисності публічного набору даних для аналітичних досліджень. *Вісник Харківського національного університету імені В. Н. Каразіна, серія Математичне моделювання. Інформаційні технології. Автоматизовані системи управління*. 2024. вип. 61. С.61-67. https://doi.org/10.26565/2304-6201-2024-61-07

How to quote: Podoliaka O. O., Podoliaka O. M., “Assessing the utility of a public dataset for analytical research”, *Bulletin of V. N. Karazin Kharkiv National University, series Mathematical modelling. Information technology. Automated control systems*, vol. 61, pp.61-67, 2024. [In Ukrainian]. https://doi.org/10.26565/2304-6201-2024-61-07

Вступ

Публічні дані містять чутливу для людей інформацію, яка може бути використана для суспільно значущих цілей, наприклад, аналітичними агентствами для складання прогнозів, керівництвом компаній для організації ефективного управління, фінансовими організаціями для визначення оптимальних бізнес-стратегій, моделювання процесів, аналізу вразливостей тощо. Важливо розуміти, що публічні дані можуть бути використані зловмисниками для злочинних цілей, таких як: шпигунство, шантаж, переслідування, здирство тощо.

Тому персональні дані мають бути деідентифіковані у публічних наборах даних. Для цього, зазвичай, застосовується шифрування, додавання статистичних шумів, узагальнення даних із

використанням таксонномій, зменшення гранулярності (точності оцінок) тощо. Очевидно, що методи деідентифікації знижують корисність даних для аналізу [1-4]. Детальні огляди методів деідентифікації можна отримати з [5-7]. Слід зазначити, що жоден з цих методів не дає повної безпеки ідентифікації людей у публічному наборі даних.

1 Ідентифікатори персональних даних та побудова публічного набору даних

Розкриття особистостей людей у «знеособлених» даних називається реідентифікацією або деанонімізацією. Реідентифікація розкриває конфіденційні дані у деідинтифікованих наборах публічних даних. В індустрії захисту персональних даних виділяють наступні типи ідентифікаторів.

Прямі ідентифікатори або явні ідентифікатори – атрибути даних, які можуть безпосередньо ідентифікувати особу. Наприклад: ID, паспорт та номер соціального страхування, ім'я – прізвище тощо. Згідно з прийнятими регуляціями, явні ідентифікатори повинні бути видалені з публічного набору даних [8-10].

Непрямі ідентифікатори або квазіідентифікатори – загальнодоступні атрибути даних, які безпосередньо не можуть ідентифікувати людину, але можуть використовуватися для ідентифікації людей у різних моделях атак. Наприклад, вік, поштовий індекс, дата народження тощо.

Конфіденційними або сенситивними ідентифікаторами називатимемо підмножину непрямих ідентифікаторів, що містять важливу особисту чи конфіденційну інформацію. Вони мають велику цінність як для зловмисників, так і для дослідників. Ці дані несуть безпосередню загрозу приватності та, у разі розкриття, можуть завдати відчутної шкоди. Наприклад, дані про здоров’я, зайнятість, освіту, матеріальне становище, релігію, різного роду уподобання тощо.

Ключовими називаються ідентифікатори, які зберігають ключі, що зв’язують таблиці публічного набору даних або посилання на інші дані (документи, фото, відео тощо).

Регламент GDPR (General Data Protection Regulation) зобов'язує видавця запобігти всіляким ризикам розкриття персональних даних у відкритих наборах даних. Основні ризики пов’язані з легкою доступністю для зловмисників обсягів персональних даних, достатніх для ідентифікації людей у знеособлених публічних даних. Хакери також можуть використовувати техніки аналізу даних для визначення кореляцій, що несуть загрозу приватності.

Початковим етапом побудови безпечного публічного набору є розробка сенситивної моделі чи моделі конфіденційних даних. Цю модель формують:

1) підмножини прямих, непрямих та конфіденційних ідентифікаторів набору даних;

2) модель конфіденційності, яка визначає критерії безпеки конфіденційних даних;

3) модель корисності даних, яка встановлює обмеження втрати інформації в ході деідентифікації набору даних.

Обов’язковим етапом побудови публічного набору даних, згідно з усіма прийнятими у світі регуляціями, є анонімізація прямих ідентифікаторів. National Institute of Standards and Technology (NIST) визначає анонімізацію, як процес, який видаляє зв’язок між ідентифікуючим набором даних та суб’єктом даних [11]. Зазвичай, анонімізація полягає у видаленні прямих ідентифікаторів (телефон, ідентифікаційний код, адреса тощо), або заміщенні їх синтетичними даними, що обчислюються згідно деякого алгоритму.

На наступному етапі захисту даних необхідно зменшити точність значень або грануляцію непрямих ідентифікаторів. Це робиться, щоб завадити зловмиснику реідентифікувати людей, коли йому відомі точні значення даних.

Слід зазначити, що існують моделі атак проти приватності, що не потребують реідентифікації, а саме, моделі журналіста та маркетолога [12, 13]. Зокрема, мета журналіста – зіпсувати репутацію видавця даних, а мета маркетолога – адекватне маркетингове дослідження. У роботі [14] розглянуто моделі атак, що ґрунтуються на розкритті значень непрямих ідентифікаторів, а також моделі оцінки ризиків відповідних загроз.

Існують дві основні техніки для зменшення точності оцінок ідентифікаторів або грануляції. Перша – додавання до даних випадкових шумів, а друга – узагальнення значень на основі ієрархій чи таксономій. Наприклад, маскування даних зірками або синтетичними значеннями можна вважати екстремальною формою узагальнення.

Зменшення грануляції можна розглядати як процедуру усунення відмінностей схожих квазіідентифікаторів. Можна сказати, що дана процедура виконує розбиття вихідної таблиці на

кластери шляхом об'єднання схожих записів (наприклад, близьких за віком, вагою, поштовим кодом тощо). Ці кластери також називають класами еквівалентності. Кожному класу відповідає множина конфіденційних даних. Ця стратегія називається «сховатися в натовпі». Під натовпом в даному випадку розуміється множина невиразних об'єктів, кожен з яких ховає свої таємниці в цьому натовпі.

2 Оцінка корисності даних або інформаційних втрат

Для оцінки корисності даних у ході деідентифікації пропонується використовувати математичні моделі теорії ймовірностей та теорії інформації Шеннона [15, 16]. Ці моделі тісно взаємопов'язані. Можна сміливо сказати, що фундаментом теорії інформації є теорія ймовірностей.

Для початку розглянемо загальну модель оцінки корисності, на підставі якої буде побудовано прикладну математичну модель оцінки корисності окремих кластерів.

Нехай джерело повідомлень X – це множина значень певного набору даних. Тоді кількість інформації окремого повідомлення $x, x \in X$ обчислюється за формулою Хартлі.

$$I(x) = -\log p(x) \tag{1}$$

де $I(x)$ показує скільки біт інформації несе повідомлення x, ймовірність якого становить $p(x)$.

Ентропія $H(X)$ – це середня кількість інформації, що надходить на одне випадкове повідомлення з джерела повідомлення X.

$$H(X) = -\sum_{x \in X} p(x) \cdot log\big(p(x)\big) \tag{2}$$

Якщо є залежність між повідомленнями або елементами множин $x \in X$, $y \in Y$, то спільна кількість інформації визначається за формулою.

$$I(x, y) = -\log p(x, y) \tag{3}$$

$$p(x, y) = p(x) \cdot p(x/y) \tag{4}$$

де $p(x, y)$ і $p(x/y)$ – спільна ймовірність і умовна ймовірність подій x, y, $p(x, y) \le p(x/y)$, $p(x, y) \le p(x)$.

Нехай X та Y – множини значень ідентифікаторів деякого публічного набору даних. Маємо на увазі, хоча це не обов'язково, що X – квазіідентифікатор, а Y – конфіденційний ідентифікатор (наприклад, X – вік пацієнта, а Y – його захворювання). Встановити наявність залежності між X та Y можна оцінивши обсяг спільної інформації цих множин $I(X, Y)$. Вона обчислюється на основі безумовної $H(X)$ та умовної ентропії $H(X/Y)$.

Припустимо, що є залежність між елементами множин $x \in X$ та $y \in Y$ деякого кластера набору даних. Тоді спільна ентропія множин X, Y визначається за формулою.

$$H(X, Y) = -\sum_{\substack{x \in X \\ y \in Y}} \big[p(x, y) \cdot log\big(p(x, y)\big)\big] \tag{5}$$

$H(X, Y)$ також називають ентропією об'єднання двох джерел, вона показує середню кількість інформації, що припадає на два випадкові повідомлення (x, y) джерел X, Y.

$$H(X, Y) = H(Y, X) = H(X) + H(Y/X) = H(Y) + H(X/Y) \tag{6}$$

$H(X/Y)$ – загальна або повна умовна ентропія об'єднання X та Y, або кількість інформації джерела X без урахування спільної інформації, що міститься в X та Y.

$$H(X/Y) = -\sum_{\substack{x \in X \\ y \in Y}} \big[p(x, y) \cdot log\big(p(x/y)\big)\big] \tag{7}$$

Загальна умовна ентропія використовується для обчислення втрат інформації джерела X з урахуванням залежності джерел X та Y.

Спільна інформація джерел X то Yвизначається за формулою.

$$I(X,Y) = I(Y,X) = H(X) - H(X/Y) = H(Y) - H(Y/X) \quad (8)$$

На рисунку 1 показано взаємозв'язок спільної ентропії, умовної ентропії та спільної інформації.

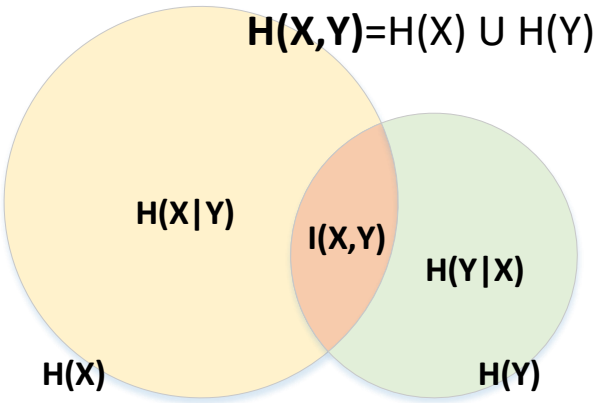


Рис. 1 Спільна, умовна ентропія та спільна інформація

Кореляцію множин X, Y можна визначити, як відношення обсягів інформації $H(X/Y)$ та $H(X)$, $0 < H(X/Y) \leq H(X)$.

$$R(X/Y) = 1 - \frac{H(X/Y)}{H(X)} \quad (9)$$

$$R(Y/X) = 1 - \frac{H(Y/X)}{H(Y)} \quad (10)$$

Область визначення R перебуває у діапазоні $[0; 1]$, тобто R відображає залежність підмножин даних.

Розглянемо переваги і недоліки моделей (6 – 8).

До переваг розглянутих моделей корисності можна віднести простоту інтерпретації оцінок, отриманих згідно з формулами (1-10). Тобто, легко зрозуміти скільки порядків точності даних (бітів) було втрачено після деідентифікації. Однак, ентропійні метрики, обчислені на основі великих обсягів даних, втрачають адекватність через екстремальне усереднення інтегральних результатів та їх обчислення є важким завданням. Ми розглянули лише два домени таблиці (X, Y). Якщо розглядати загальний випадок, коли $X \epsilon X^*; Y \epsilon Y^*$, де X^*– множина всіх квазіідентифікаторів, а Y^*– множина всіх конфіденційних ідентифікаторів у формули (6-8) слід додати суми за цими множинами. Оскільки обчислення звичайних ймовірностей має лінійну складність, а умовних – квадратичну, для таблиць з мільйонами записів і сотнями доменів обчислення розглянутих оцінок перетворюється на нездійсненну обчислювальну задачу.

Для вирішення описаних вище проблем пропонується два підходи.

1. Формування менш точного набору даних шляхом зменшення грануляції або точності метричних оцінок атрибутів початкових даних.

2. Формування випадкової репрезентативної вибірки даних задля підвищення швидкості обчислення необхідних статистик.

Ці підходи вирішують проблеми обчислюваності ентропійних метрик великих баз даних.

Перший підхід. Неточний тимчасовий набір даних необхідний виключно, щоб швидко порахувати необхідні ймовірності для метрик корисності. Важливо розуміти, що для аналізу здебільшого не потрібна екстремальна точність. Наприклад, якщо база містить історії транзакцій за рахунками клієнтів, то навіщо маркетологу знати історії платежів до останнього долара або

цента. Унікальність значень ідентифікаторів – це пряма загроза приватності. Наприклад, за сумою платежу в знайденому чеку хакер може з великою ймовірністю ідентифікувати особистість в базі, з якої видалені прямі ідентифікатори.

Другий підхід. Слід підкреслити, що в теорії інформації обчислення розглянутих ентропій виконується в повній групі подій. Це означає, що необхідні ймовірності мають обчислюватися по всій базі даних. Оскільки отримання необхідних статистик у повній групі подій неможливе, пропонується обчислити необхідні ймовірності на підставі репрезентативної вибірки невеликого розміру. У базі даних записи зазвичай упорядковані за деяким індексом, тому отримання випадкової вибірки з великої бази являє собою складну обчислювальну задачу. Ефективні методи отримання випадкової вибірки розглянуті в роботах [17-19].

Важливо розуміти, що на вході кластеризації значення даних різних рядків таблиці спотворюються не рівномірно. Отже, втрати інформації окремих елементів даних і кластерів публічного набору даних можуть відрізнятися. Тому потрібна метрика корисності для окремого кластера та груп кластерів. Ентропійні метрики кластера, обчислені на основі статистики всієї таблиці неможливо застосувати для кластерів, оскільки ентропії мають обчислюватися для повної групи подій. Тому, для оцінки втрат інформації або корисності окремого кластеру чи підмножин кластерів пропонується використовувати відношення кількості інформації деідентифікованого кластера до кількості інформації вихідної підмножини даних. Дану метрику легко порахувати, її також можна використовувати в якості одного з критеріїв кластеризації у процесі побудови публічного набору даних.

Висновки.

Описано процедури побудови безпечного публічного набору даних. Розглянуто питання застосування ентропійних метрик теорії інформації Шеннона для оцінки втрат інформації та корисності даних для аналізу. Показано, що обчислення зазначених метрик є складною, практично не здійсненною для великих баз даних, обчислювальною задачею. Запропоновано процедури, спрямовані на підвищення швидкодії обчислень розглянутих метрик. А саме, створення менш точної копії вихідних даних та формування випадкової вибірки із великої бази даних для обчислення необхідних статистик. Розглянуто метрики оцінки корисності для окремих підмножин (кластерів) публічних даних.

СПИСОК ЛІТЕРАТУРИ

1. Dankar, F., Emam, K.E., Neisa, A., Roffey, T.: Estimating the re-identification risk of clinical data sets. BMC Med. Inform. Decis. Mak, 2012. №12 (66). С. 1-15.
2. B.A. Malin, D. Karp, R.H. Scheuermann. Technical and policy approaches to balancing patient privacy and data sharing in clinical and translational research. J. Investig. Med., 2010. 58 (1). С. 11-18.
3. Li, Tiancheng & Li, Ninghui. On the tradeoff between privacy and utility in data publishing. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2009. С. 517-526.
4. Fung, Benjamin & Wang, ke & Chen, Rui & Yu, Philip. Privacy-Preserving Data Publishing: A Survey of Recent Developments. ACM Comput. Surv, 2010. №4 (14). С. 1-53.
5. Yaseen, Saba & Abbas, Syed & Anjum, Adeel & Saba, Tanzila & Khan, Abid & Malik, Saif & Ahmad, Naveed & Shahzad, Basit & Bashir, Ali. Improved Generalization for Secure Data Publishing. IEEE Access, 2018. С. 27156-27165.
6. Fung, Benjamin & Wang, Ke & Fu, Ada & Yu, Philip. Introduction to Privacy-Preserving Data Publishing: Concepts and Techniques, 2010. 376 с. ISBN: 9780429138737.
7. Li, Ninghui & Li, Tiancheng & Venkatasubramanian, Suresh. t-Closeness: Privacy Beyond k-Anonymity and l-Diversity. IEEE 23rd International Conference on Data Engineering (ICDE), 2007. 2. С. 106 - 115.
8. Sweeney, L.: Achieving k-anonymity privacy protection using generalization and suppression. J. Uncertain. Fuzz. Knowl. Sys., 2002. 10 (5), С. 571-588.
9. US Department of Health and Human Services. Guidance regarding methods for de-identification of protected health information in accordance with the health insurance portability and accountability act (HIPAA) privacy rule, 2014. [Електронний ресурс] / Режим доступу: http://www.hhs.gov/.

10. Simson L. Garfinkel. NISTIR 8053. De-Identification of Personal Information, 2015. [Електронний ресурс] / Режим доступу: http://dx.doi.org/10.6028/NIST.IR.8053
11. Fung B., Wang ke, Wang L., Debbabi M. A framework for privacy-preserving cluster analysis. Conference: Intelligence and Security Informatics, 2008. С. 46 - 51.
12. Emam K., Dankar F. (2008). Protecting Privacy Using k-Anonymity. Journal of the American Medical Informatics Association : JAMIA. 2008. 15(5).
13. Marques, Joana & Bernardino, Jorge. Analysis of Data Anonymization Techniques. In Proceedings of the 12th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management, 2020. С. 235-241. ISBN: 978-989-758-474-9.
14. Podoliaka O., Mushkatblat V., Kaplan A. Privacy Attacks Based on Correlation of Dataset Identifiers: Assessing the Risk, 2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC), 2022. С. 0808-0815. ISBN: 9781665483032.
15. Шеннон К. Работы по теории информации и кибернетике. Издательство иностранной литературы. 1963. 830 с.
16. Шнайер, Б. Секреты и ложь. Безопасность данных в цифровом мире. – Пер. с англ. – СПб.: Питер, 2004. с. 432. ISBN: 5-318-00193-9.
17. Быстрый выбор случайных значений из больших таблиц MySQL по условию. [Електронний ресурс] / Доступно: https://habr.com/ru/post/207096/
Дата звернення: Трав. 1, 2022.
18. Greg Robidoux. Retrieving random data from SQL Server with TABLESAMPLE. [Електронний ресурс] / Доступно: https://www.mssqltips.com/sqlservertip/1308/retrieving-random-data-from-sql-server-with-tablesample/.
Дата звернення: Трав. 1, 2022.
19. NOTES ON SQL. [Електронний ресурс] / Доступно: https://sqlrambling/.net/2018/01/24/tablesample-basic-examples.
Дата звернення: Трав. 1, 2022.

REFERECES

1. Dankar, F., Emam, K.E., Neisa, A., Roffey, T.: Estimating the re-identification risk of clinical data sets. BMC Med. Inform. Decis. Mak, 2012. №12 (66). P. 1-15.
2. B.A. Malin, D. Karp, R.H. Scheuermann. Technical and policy approaches to balancing patient privacy and data sharing in clinical and translational research. J. Investig. Med., 2010. 58 (1). P. 11-18.
3. Li, Tiancheng & Li, Ninghui. On the tradeoff between privacy and utility in data publishing. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2009. P. 517-526.
4. Fung, Benjamin & Wang, ke & Chen, Rui & Yu, Philip. Privacy-Preserving Data Publishing: A Survey of Recent Developments. ACM Comput. Surv, 2010. №4 (14). P. 1-53.
5. Yaseen, Saba & Abbas, Syed & Anjum, Adeel & Saba, Tanzila & Khan, Abid & Malik, Saif & Ahmad, Naveed & Shahzad, Basit & Bashir, Ali. Improved Generalization for Secure Data Publishing. IEEE Access, 2018. P. 27156-27165.
6. Fung, Benjamin & Wang, Ke & Fu, Ada & Yu, Philip. Introduction to Privacy-Preserving Data Publishing: Concepts and Techniques, 2010. 376 s. ISBN: 9780429138737.
7. Li, Ninghui & Li, Tiancheng & Venkatasubramanian, Suresh. t-Closeness: Privacy Beyond k-Anonymity and l-Diversity. IEEE 23rd International Conference on Data Engineering (ICDE), 2007. 2. P. 106 - 115.
8. Sweeney, L.: Achieving k-anonymity privacy protection using generalization and suppression. J. Uncertain. Fuzz. Knowl. Sys., 2002. 10 (5). P. 571-588.
9. US Department of Health and Human Services. Guidance regarding methods for de-identification of protected health information in accordance with the health insurance portability and accountability act (HIPAA) privacy rule, 2014. available at:http://www.hhs.gov/.
10. Simson L. Garfinkel. NISTIR 8053. De-Identification of Personal Information, 2015. available at: http://dx.doi.org/10.6028/NIST.IR.8053.
11. Fung B., Wang ke, Wang L., Debbabi M. A framework for privacy-preserving cluster analysis. Conference: Intelligence and Security Informatics, 2008. P. 46 - 51.

12. Emam K., Dankar F. (2008). Protecting Privacy Using k-Anonymity. Journal of the American Medical Informatics Association : JAMIA. 2008. 15(5).
13. Marques, Joana & Bernardino, Jorge. Analysis of Data Anonymization Techniques. In Proceedings of the 12th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management, 2020. P. 235-241. ISBN: 978-989-758-474-9.
14. Podoliaka O., Mushkatblat V., Kaplan A. Privacy Attacks Based on Correlation of Dataset Identifiers: Assessing the Risk, 2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC), 2022. P. 0808-0815. ISBN: 9781665483032.
15. Shennon K. Raboty po teorii informacii i kibernetike. Izdatel'stvo inostrannoj literatury, 1963. 830 s.
16. Shnajer, B. Sekrety i lozh'. Bezopasnost' dannyh v cifrovom mire. – Per. s angl. – SPb.: Piter, 2004. s. 432. ISBN: 5-318-00193-9.
17. Bystryj vybor sluchajnyh znachenij iz bol'shih tablic MySQL po usloviyu. available at: https://habr.com/ru/post/207096/.
Available: May. 1, 2022.
18. Greg Robidoux. Retrieving random data from SQL Server with TABLESAMPLE. available at: https://www.mssqltips.com/sqlservertip/1308/retrieving-random-data-from-sql-server-with-tablesample/.
Available: May. 1, 2022.
19. NOTES ON SQL. available at: https://sqlrambling.net/2018/01/24/tablesample-basic-examples.
Available: May. 1, 2022.

Podoliaka Oksana *PhD of Technical Sciences, docent*
V.N. Karazin Kharkiv National University, Svobody Square 4, Kharkiv, Ukraine, 61022
Podoliaka Oleksii *Senior lecturer*
V.N. Karazin Kharkiv National University, Svobody Square 4, Kharkiv, Ukraine, 61022

Assessing the utility of a public dataset for analytical research

Organizations and agencies release various data intended for analysis, training of artificial intelligence systems, and other research purposes. According to the adopted regulations in the field of personal data protection, public data must be anonymized and protected from various threats of personal data disclosure. Elimination of these threats is realized by reducing the accuracy of data during their preparation for the release. Loss of accuracy obviously leads to a decrease in the usefulness of data for analysis. The paper considers entropy metrics of utility and problems of their computability, as well as metrics of loss of utility of certain subsets of public data.

Objective. To develop effective metrics for assessing the usefulness of a public dataset for analysis, taking into account the requirements of personal data protection.

Research methods. Information security, Shannon's theory of information, Data Governance.

Results. Metrics for assessing information loss and data usefulness for analysis based on the entropy metrics of Shannon's information theory are proposed. Procedures aimed at increasing the speed of calculations of the considered metrics are suggested.

Conclusions. The procedures for building a secure public dataset are described. The application of entropy metrics of Shannon's information theory to assess information loss and data usefulness for analysis is considered. It has been shown that the calculation of these metrics is a complex computational task that is practically impossible for large databases. Procedures aimed at increasing the speed of calculating the considered metrics are proposed. In particular, the creation of a less accurate copy of the original data and the formation of a random sample from a large database to calculate the necessary statistics. The metrics for assessing the usefulness of certain subsets (clusters) of public data are considered in the article.

Keywords: *data privacy, de-identification, data publishing, data utility, GDPR (General Data Protection Regulation).*

Надійшла у першій редакції 04.03.2024, в останній - 01.05.2024.
The first version has been received on 04.03.2024, the final version - on 01.05.2024.

Наукове видання

Вісник Харківського національного університету імені В. Н. Каразіна

Серія «Математичне моделювання. Інформаційні технології. Автоматизовані системи управління»

Випуск 61

Збірник наукових праць

Українською та англійською мовами

Комп'ютерне верстання О. О. Афанасьєва

Підписано до друку 15.12.2023 р.
Формат 60x84/8. Папір офсетний. Друк цифровий.
Ум. друк. арк. – 6,67.
Обл.– вид. арк. – 8,33.
Наклад 50 пр. Зам. № 55/2024
Безкоштовно

Видавець і виготовлювач
Харківський національний університет імені В. Н. Каразіна
61022, м. Харків, майдан Свободи, 4
Свідоцтво суб'єкта видавничої справи ДК №3367 від 13.01.09

Видавництво Харківський національний університет імені В. Н. Каразіна
тел.: 705-24-32