

СУЧАСНІ ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

№ 1 (1) 2021

НАУКОВИЙ ЖУРНАЛ

ISSN 2788-6603

КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ТАРАСА ШЕВЧЕНКА

Подано теоретичні та практичні результати досліджень із програмної інженерії, прикладних інформаційних систем і технологій, мережних та інтернет-технологій, інженерії знань, штучного й обчислювального інтелекту, математичних основ інформаційних технологій, інформаційної аналітики й аналітики даних, машинного навчання та розпізнавання образів, цифрових технологій управління IT-проектами, розумних систем та інтернету речей.

Для науковців, фахівців, студентів.

ГОЛОВНИЙ РЕДАКТОР	В. Є. Снитюк , д-р техн. наук, проф.
ЗАСТУПНИК ГОЛОВНОГО РЕДАКТОРА	К. В. Колесникова , д-р техн. наук, проф.
ВІДПОВІДАЛЬНІ СЕКРЕТАРІ	Н. П. Тмєнова , канд. фіз.-мат. наук, доц. О. В. Федусенко , канд. техн. наук, доц.
РЕДАКЦІЙНА КОЛЕГІЯ	Алі Аль-Амморі, О. В. Барабаш, О. В. Бісікало, О. С. Бичков, Є. В. Гаврилко, Г. М. Гнатієнко, М. М. Корабльов, Я. І. Корнага, Ю. В. Кравченко, Є. В. Малахов, О. О. Марченко, В. Машков, О. М. Медведєва, В. В. Морозов, В. Є. Мухін, В. Л. Плєскач, О. Г. Плющ, Ю. Я. Самохвалов, А. О. Семенов, С. Ф. Теленик, І. Б. Туркін, В. В. Циганок, Дархан Ахмед-Закі, Хуберт Рот, В. Вовк
Адреса редколегії	вул. Богдана Гаврилишина, 24, Київ, 04116, Україна ☎ факс (+38044) 481 44 07, e-mail: fit@univ.net.ua
Затверджено	Вченою радою факультету інформаційних технологій Київського національного університету імені Тараса Шевченка 21.12.21 (протокол № 5)
Зареєстровано	Міністерством інформації України. Свідоцтво про державну реєстрацію КВ № 24719-14659Р від 04.01.2021
Засновник та видавець	Київський національний університет імені Тараса Шевченка, ВПЦ "Київський університет" Свідоцтво внесено до Державного реєстру ДК № 1103 від 31.10.02
Адреса видавця	ВПЦ "Київський університет", б-р Тараса Шевченка, 14, м. Київ, 01601, Україна ☎ (+38044) 239 31 72, 239 32 22; факс 239 31 28

ADVANCED INFORMATION TECHNOLOGY

№ 1 (1) 2021

SCIENTIFIC JOURNAL

ISSN 2788-6603

NATIONAL TARAS SHEVCHENKO UNIVERSITY OF KYIV

Theoretical and practical results of research in software engineering, applied information systems and technologies, network and Internet technologies, knowledge engineering, artificial and computational intelligence, mathematical foundations of information technology, information analytics and data analytics, machine learning and pattern recognition, digital management technologies of IT projects, smart systems and the Internet of Things are presented.

For scientists, specialists, students.

EDITOR-IN-CHIEF	Vitaliy Snytyuk , Dr Hab., Prof.
DEPUTY OF EDITOR-IN-CHIEF	Kateryna Kolesnikova , Dr Hab., Prof.
EXECUTIVE SECRETARIES	Tmienova Nataliia , Phd, Associate Prof. Olena Fedusenko , Phd, Associate Prof.
EDITORIAL BOARD	Ali Al-Ammouri, Oleg Barabash, Oleg Bisikalo, Olexiy Bychkov, Yevhen Havrylko, Hrygorii Hnatiienko, Mykola Korablev, Yuriy Kravchenko, Yaroslav Kornaga, Yuriy Kravchenko, Eugene Malakhov, Oleksandr Marchenko, Viktor Mashkov, Olena Medvedieva, Victor Morozov, Vadym Mukhin, Valentyna Pleskach, Olexandr Pliushch, Yuriy Samokhvalov, Andriy Semenov, Sergii Telenyk, Igor Turkin, Vitaliy Tsyganok, Darkhan Ahmed-Zaki, Hubert Roth, Vladimir Vovk
Editorial board address	Bohdan Hawrylyshyn str. 24, UA-04116, Kyiv, Ukraine ☎ fax (+38044) 481 44 07, e-mail: fit@univ.net.ua
Confirmed	Academic Council of the Faculty of Information Technology of Taras Shevchenko National University of Kyiv 21.12.21 (protocol № 5)
Registered	Ministry of Information of Ukraine. State registration certificate KB № 24719-14659P from 04.01.2021
Founder and Publisher	Taras Shevchenko National University of Kyiv, Publishing and Polygraphic Center "Kyiv University" The certificate is added to registry ДК № 1103 from 31.10.02
Publisher's address	Publishing and Polygraphic Center "Kyiv University" 14, Taras Shevchenko blvd., Kiev, 01601, Ukraine ☎ (38044) 239 3172, 239 32 22; fax 239 31 28

Прикладні інформаційні системи та технології

Оксана Мулеса, Віталій Снитюк

Клієнтоорієнтована оптимізація кадрового складу
закладів охорони здоров'я..... 6

*Олена Федусенко, Наталія Шкурпела,
Ірина Доманецька, Анатолій Федусенко*

Інтелектуальна система підтримки прийняття
агротехнологічних рішень для посіву полів..... 15

Штучний та обчислювальний інтелекти

Марина Антоневиц, Анна Дідик, Наталія Тмєнова, Віталій Снитюк

Вибір найкращих параметрів для методу
деформованих зірок у n-вимірному просторі..... 24

Математичні основи інформаційних технологій

Микола Пирог, Ірина Гарко, Ксенія Духновська

Моделювання термofізіологічного стану людини..... 30

Григорій Гнатієнко, Олексій Олецький

Порівняння методик ранжування науковців на основі
індексів Гірша та PageRank: суперечливі ситуації 39

Машинне навчання та розпізнавання образів

Олег Іларіонов, Антон Астахов, Ганна Красовська, Ірина Доманецька

Інтелектуальний модуль розпізнавання емоцій за голосом..... 46

Ірина Юрчук, Олена Колесник

Сегментація як ефективний метод виокремлення
пухлини головного мозку на зображенні МРТ 53

Микола Мороз, Денис Берестов, Олег Курченко

Аналіз алгоритмів візуального відстежування об'єктів
для систем реального часу 59

Інформаційна аналітика й аналітика даних

Людмила Зубик, Ярослав Зубик

Архітектура сучасних платформ для аналітики великих даних..... 67

Цифрові технології управління ІТ-проектами

Віктор Морозов, Максим Проскурін

Моделі проактивного управління змінами у проектах
створення та розвитку розподілених інформаційних систем 76

Олександр Тімінський, Анна Коломієць, Ольга Мезенцева

Моделі управління проектом створення
ІТ-компанії у галузі Data Science..... 86

Applied information systems and technology*Oksana Mulesa, Vitaliy Snytyuk*

Customer-oriented optimization of healthcare facilities staff..... 6

Olena Fedusenko, Natalia Shkurpela, Iryna Domanetska, Anatoliy Fedusenko

Intelligent support system for agro-technological decision-making for field sowing 15

Artificial and computational intelligence*Maryna Antonevych, Anna Didyk, Nataliia Tmienova, Vitaliy Snytyuk*

Choosing the best parameters for method of deformed stars in n-dimensional space 24

Mathematical foundations of information technology*Mykola Pyroh, Iryna Harko, Ksenia Dukhnovska*

Modeling of thermophysiological state of person..... 30

*Hryhorii Hnatiienko, Oleksiy Oletsky*Comparing methods of ranking scientists on the base of Hirsch index
and PageRank: contradictory situations 39**Machine learning and pattern recognition***Oleh Ilarionov, Anton Astakhov, Hanna Krasovska, Iryna Domanetska*

Intelligent module for recognizing emotions by voice 46

Iryna Yurchuk, Olena Kolesnyk

Segmentation as an effective method of isolating a brain tumor on MRI 53

Mykola Moroz, Denis Berestov, Oleg Kurchenko

Analysis of visual object tracking algorithms for real-time systems 59

Information analytics and data analytics*Liudmyla Zubyk, Yaroslav Zubyk*

Architecture of modern platforms for big data analytics 67

Digital technologies of IT project management*Viktor Morozov, Maksym Proskurin*Models of proactive change management in projects for the creation
and development of distributed information systems 76*Alexander Timinsky, Anna Kolomiets, Olga Mezentseva*

Project management models for creating an IT company in the field of Data Science 86



ПРИКЛАДНІ ІНФОРМАЦІЙНІ СИСТЕМИ ТА ТЕХНОЛОГІЇ



УДК 004.89

DOI: <https://doi.org/10.17721/AIT.2021.1.01>

О. Ю. Мулеса,
orcid.org/0000-0002-6117-5846,
ДВНЗ «Ужгородський національний університет», Ужгород, Україна,
В. Є. Снитюк,
orcid.org/0000-0002-9954-8767,
Київський національний університет імені Тараса Шевченка, Київ, Україна

Клієнтоорієнтована оптимізація кадрового складу закладів охорони здоров'я

Розглянуто проблему розроблення інформаційної технології для клієнтоорієнтованої оптимізації кадрового складу закладів охорони здоров'я. Установлено, що ключовими для напрацювання управлінських рішень у процесах формування кадрової політики в медичних закладах є задачі оцінювання таких характеристик кадрового складу: нормативне та фактичне навантаження на працівників, задачі прогнозування потреб у медичних послугах у майбутньому. Їхнє розв'язання дозволить одержати та використати нові знання. Відповідний процес базується на послідовному розв'язанні задач кластеризації, ідентифікації та прогнозування. Побудовано математичну модель задачі багатокритеріальної оптимізації для формування множини варіантів рішень щодо формування кадрової політики в закладах охорони здоров'я. Виконано модель пошуку оптимального розподілу послуг між працівниками закладу у формі задачі цілочислового лінійного програмування. Розроблено продукційні правила для аналізу результатів розв'язування сформульованої задачі. Запропоновано модель для напрацювання варіантів рішень щодо зміни в кадровому складі медичного закладу для забезпечення повноти надання медичних послуг. Для оцінювання ефективності варіантів вироблених рішень використовується індекс клієнтоорієнтованості, який обчислюється відповідно до принципів егалітаризму. Виконано експериментальну верифікацію розроблених моделей і методів.

Ключові слова: кадровий потенціал, клієнтоорієнтована оптимізація, індекс клієнтоорієнтованості закладу, прийняття рішень, оптимізація.

Для цитування (for citation): О. Ю. Мулеса, В. Є. Снитюк "Клієнтоорієнтована оптимізація кадрового складу закладів охорони здоров'я", *Сучасні інформаційні технології*, vol.1, pp. 6–14, 2021.

ВСТУП

Процеси проектування та моделювання життєвого циклу складних соціальних систем маємо під час розв'язання проблем, пов'язаних з прийняттям рішень щодо організації функціонування підприємств та установ у різних сферах життєдіяльності людини [1]. Метою таких процесів, серед інших, є розроблення моделей і методів для ідентифікації значень параметрів системи, які забезпечили б її ефективно, за деяким заданим критерієм, функціонування.

Заклади охорони здоров'я (ЗОЗ) структурно та функціонально є складними системами [2]. Їхню діяльність розглядають у декількох аспектах, ключовим з яких є надання медичних послуг населенню. Визначальну роль у цих процесах відіграє кадровий склад закладу. Права та функціональні обов'язки медичних і немедичних працівни-

ків ЗОЗ залежать від таких показників: фах, категорія, посада, оклад тощо. У ході оцінювання ефективності кадрового складу ЗОЗ ключовим є поняття якості наданих медичних послуг. Важливо також забезпечити своєчасність і повноту наданих медичних послуг, за якими до закладу звернулися пацієнти.

Медичні заклади функціонують в умовах невизначеності [3]. У випадку прийняття рішень щодо формування кадрового складу для виконання ситуаційного аналізу та прогнозування майбутніх потреб у медичних послугах необхідно обробляти й аналізувати дані різної природи. Для оцінювання основних параметрів діяльності кадрового складу ЗОЗ розв'язують задачі препроцесінгу даних, кластеризації, класифікації, ідентифікації, прогнозування тощо. Результати розв'язання таких задач та одержані знання можуть стати пере-

© Мулеса О. Ю., Снитюк В. Є., 2021



думовою для напрацювання та прийняття ефективних управлінських рішень щодо формування кадрового складу ЗОЗ на майбутні періоди часу.

АНАЛІЗ ЛІТЕРАТУРНИХ ДАНИХ І ПОСТАНОВКА ЗАДАЧІ

Розглядається проблема клієнтоорієнтованої оптимізації кадрового складу ЗОЗ, а саме необхідність знаходження таких варіантів рішень, реалізація яких дозволить сформувати кадровий склад, необхідний для забезпечення попиту на медичні послуги клієнтів закладу. Формування кадрового складу пропонується виконувати на основі аналізу поточних показників, які характеризують навантаження на працівників закладу, а також з урахуванням прогностичних значень щодо попиту на медичні послуги в майбутні періоди часу.

Проблемам оцінювання основних показників кадрового складу ЗОЗ присвячено роботи [4, 5]. В основу розроблених моделей і методів покладено ідею врахування фотохронометражів медичних послуг [6, 7], дотримання яких є одним із ключових критеріїв якості послуги. У [3] показано, що у процесах оцінювання майбутніх потреб у медичних послугах розв'язують задачі кластеризації, класифікації, ідентифікації та прогнозування.

Методам розв'язання задач кластеризації присвячено велику кількість сучасних наукових досліджень. Так, роботи [8, 9] містять аналіз статистичних методів кластеризації, [10] – деревовидних, у роботі [11] наведено нейромережні методи кластеризації тощо.

Для розв'язання задачі структурної та параметричної ідентифікації невідомих залежностей часто використовують метод групового врахування аргументів [12], який розвинено на нечіткий випадок [13, 14].

Методи прогнозування на основі часових рядів умовно поділяють на інтуїтивні (наприклад, методи на основі експертних оцінок [15], нечіткі методи прогнозування [16]) та формалізовані методи, серед яких особливої актуальності набувають гібридні методи прогнозування [17–19].

Прогностичні значення попиту на медичні послуги в майбутньому стануть підґрунтям для напрацювання варіантів управлінських рішень щодо формування кадрового складу ЗОЗ і розподілу функціональних обов'язків між працівниками закладу. Важливою залишається проблема оцінювання ефективності варіантів рішень за критерієм підвищення якості надаваних кадровим складом медичних послуг.

ПРОБЛЕМА ВИЗНАЧЕННЯ ОПТИМАЛЬНОГО НАВАНТАЖЕННЯ НА ПРАЦІВНИКІВ ЗАКЛАДУ ОХОРОНИ ЗДОРОВ'Я

Напрацювання рішень щодо розподілу навантаження між працівниками ЗОЗ доцільно виконувати на основі аналізу всіх доступних даних, які умовно можна поділити на такі групи [3]:

1. Набори даних, які містять відомості про кадровий склад закладу (W), серед яких дані про фах, посаду, спеціалізацію, перелік послуг, які вони можуть надавати, тощо.

2. Відомості про особливості функціонування закладу (S), перелік послуг, які доступні для надання клієнтам закладу.

3. Ретроспективні дані у формі динамічних рядів (Ds) та статистичних даних (St) про значення величин за минулі періоди часу.

4. Результати опитувань експертів (E) з питань, що стосуються проблеми.

5. Відомості, які отримані в результаті проведення анкетування осіб – потенційних споживачів медичних послуг (Q).

6. Колективні рішення щодо розширення, перепрофілювання, відкриття філій тощо, насамперед у приватних ЗОЗ (V).

Отже, формально задачі обчислення основних показників, які характеризують діяльність закладу, можна представити так:

1. Задача оцінювання кадрового потенціалу закладу охорони здоров'я, який визначається як сумарний обсяг послуг, які потенційно можуть бути надані працівниками закладу у межах їхнього робочого часу. Необхідно сконструювати правило Θ для того, щоб на основі заданих числових характеристик ЗОЗ, можна було б визначити нормативне навантаження на працівників закладу у заданому вигляді (NB):

$$\Theta: (W, S, E) \rightarrow NB. \quad (1)$$

2. Задача оцінювання наявного навантаження на працівників ЗОЗ [4].

Побудувати правило Δ , за яким на основі відомостей про ЗОЗ і статистичних даних про послуги, що були надані працівниками закладу за визначений період часу, можна було б визначити фактичне навантаження на працівників закладу (RB):

$$\Delta: (W, S, St, E) \rightarrow RB. \quad (2)$$

3. Задача прогнозування майбутніх потреб у медичних послугах полягає в необхідності побудови правила Ω для оцінювання рівня попиту на медичні послуги (D):



$$\Omega: (S, D_s, St, E, Q) \rightarrow D. \quad (3)$$

Розв'язання задачі (3) залежить від таких підзадач:

- прогнозування основних кількісних характеристик попиту на медичні послуги на основі ретроспективних даних;
- розбиття потенційних споживачів медичних послуг на кластери, відповідно до їхніх поведінкових особливостей та з урахуваннями характерних рис їхніх соціально-демографічних портретів;
- прогнозування виникнення потреб у медичних послугах в окремих групах потенційних споживачів послуг.

4. Задача прийняття управлінських рішень щодо діяльності ЗОЗ: побудувати правило ψ для формування множини ефективних рішень (ED) на основі даних, отриманих від використання правил (1)–(3):

$$\psi: (W, NB, RB, D, E, V) \rightarrow ED. \quad (4)$$

На основі множини ED , сформованої в результаті розв'язання задачі (4), особа, що приймає рішення, може формувати програму оптимізації кадрового складу ЗОЗ.

Під оптимізацією розуміємо процес знаходження такого розподілу послуг між працівниками закладу, який задовольняє такі критерії:

- забезпечення дотримання визначених клінічними протоколами та внутрішніми розпорядками відповідностей між послугами та працівниками, які їх надають;
- забезпечення дотримання визначених нормативно-правовими актами норм робочого часу працівників;
- дотримання хронометражів медичних послуг;
- задоволення попиту на медичні послуги тощо.

На основі сформульованих критеріїв можливою є побудова моделі задачі багатокритеріальної оптимізації таким чином. Позначимо через x_{ij} – кількість послуг із номером j , $j = \overline{1, |S|}$, які надав або планує надати працівник із номером i , $i = \overline{1, |W|}$. Сформуємо матрицю компетентностей $MQ = \{\beta_{ij}\}$, де $\beta_{ij} = 1$, якщо працівник W_i може надавати послугу S_j , та $\beta_{ij} = 0$ – у протилежному випадку.

Тоді для визначення оптимального розподілу послуг між працівниками закладу необхідно розв'язати таку багатокритеріальну задачу:

$$\sum_{i=1}^{|W|} \sum_{j=1}^{|S|} x_{ij} \beta_{ij} \rightarrow \max, \quad (5)$$

$$\sum_{i=1}^{|W|} |NB_i - RB_i| \rightarrow \min, \quad (6)$$

$$\sum_{i=1}^{|W|} \left| \sum_{j=1}^{|S|} x_{ij} s_j - RB_i \right| \rightarrow \min, \quad (7)$$

$$\sum_{j=1}^{|S|} \left| \sum_{i=1}^{|W|} x_{ij} - D_j \right| \rightarrow \min, \quad (8)$$

$$x_{ij} \in N \cup \{0\}, \quad i = \overline{1, |W|}, \quad j = \overline{1, |S|}.$$

Очевидно, що багатокритеріальна задача (5)–(8) містить конфліктуючі цільові функції, таким чином, для знаходження варіантів її розв'язків, особа, що приймає рішення, має, наприклад, провести ранжування цільових функцій або задачі додаткові обмеження.

МОДЕЛЬ ПОШУКУ ОПТИМАЛЬНОГО РОЗПОДІЛУ ПОСЛУГ МІЖ ПРАЦІВНИКАМИ ЗАКЛАДУ

У роботі [20] запропоновано моделі пошуку оптимального розподілу послуг між працівниками на основі результатів прогнозування потреб у медичних послугах у майбутні періоди часу.

Задачу пошуку оптимального розподілу послуг між працівниками на період часу T побудовано у вигляді лінійної задачі цілочислового програмування таким чином:

Знайти

$$F = \sum_{i=1}^n \sum_{j=1}^m \beta_{ij} s_j x_{ij} \rightarrow \max, \quad (9)$$

де x_{ij} – кількість послуг S_j , які надаватиме працівник W_i за таких обмежень:

$$\sum_{i=1}^N x_{ij} \geq p_{jT}, \quad j = \overline{1, M} \quad (10)$$

$$\sum_{j=1}^M s_j x_{ij} \leq r_{iT}, \quad i = \overline{1, N}, \quad (11)$$

$$x_{ij} \in Z^+, \quad i = \overline{1, N}, \quad j = \overline{1, M}, \quad (12)$$

$$x_{ij} = 0, \quad \text{якщо } \beta_{ij} = 0, \quad (13)$$

де r_{iT} – нормативне навантаження на працівника W_i за період часу T .

Під час розв'язання задачі (9)–(13) можливі такі випадки:

1. Існує оптимальний розв'язок x_{ij}^* , $i = \overline{1, N}$, $j = \overline{1, M}$ задачі, який і буде характеризувати оптимальний розподіл послуг між працівниками. Тоді оптимальне навантаження на працівників обчислюється як

$$r_{iT}^* = \sum_{j=1}^m x_{ij}^* s_j.$$



2. Якщо задача (9)–(13) має декілька оптимальних розв'язків, то можливим є вироблення декількох варіантів рішень щодо оптимального розподілу навантаження між працівниками закладу, аналогічно до першого випадку.

3. Задача (9)–(13) може не мати розв'язків у таких випадках:

- якщо $\sum_{j=1}^M p_{jT} > R_T$, тобто, коли наявних

кредитів усіх працівників закладу не вистачає, щоб надати всі необхідні послуги. У цьому випадку актуальним є або скорочення переліку послуг та їхніх обсягів, або збільшення кількості працівників закладу.

- якщо для деяких окремих послуг не вистачає кредитів працівників. У цьому разі необхідно аналізувати структури й обсяг послуг, які необхідно надати працівникам закладу, та приймати рішення або про перекваліфікацію працівників, або про розширення штату працівників закладу.

Для формування множини можливих управлінських рішень щодо залучення нових працівників або звуження кількості послуг, які надаються закладом, задачу (9)–(13) можна перевизначити так:

$$F = \sum_{i=1}^N \sum_{j=1}^M \beta_{ij} s_j x_{ij} \rightarrow \max, \quad (14)$$

$$\sum_{i=1}^N x_{ij} \geq p_{jT} - \delta p_{jT}, \quad j = \overline{1, M}, \quad (15)$$

$$\sum_{j=1}^M s_j x_{ij} \leq r_{iT} + \delta r_{iT}, \quad i = \overline{1, N}, \quad (16)$$

$$x_{ij} \in Z^+, \quad i = \overline{1, N}, \quad j = \overline{1, M}. \quad (17)$$

$$x_{ij} = 0, \quad \text{якщо } \beta_{ij} = 0, \quad (18)$$

де $\delta p_{jT} \geq 0$ – величина, яка показує, наскільки будуть зменшені обсяги послуги S_j , а δr_{iT} – допустиме значення збільшення навантаження на працівника W_i .

Задачі (14)–(18) можуть розв'язуватись для різних значень δp_{jT} та δr_{iT} і, залежно від отриманих результатів, можливим є прийняття відповідних управлінських рішень. Наприклад, якщо для деякого працівника W_i $\delta r_{iT} \approx r_{iT}$, а відповідне йому Δ_{iT} достатньо невелике, то звідси слідує, що для надання послуг закладу необхідно залучити ще одного працівника з такими самими посадовими функціями та характеристиками, як і працівник W_i .

ОЦІНЮВАННЯ ЕФЕКТИВНОСТІ ВАРІАНТІВ УПРАВЛІНСЬКИХ РІШЕНЬ ЩОДО КЛІЄНТООРІЄНТОВАНОЇ ОПТИМІЗАЦІЇ КАДРОВОГО СКЛАДУ ЗАКЛАДІВ ОХОРОНИ ЗДОРОВ'Я

Для оцінювання ефективності варіантів вироблених управлінських рішень щодо клієнтоорієнтованої оптимізації кадрового складу ЗОЗ уведемо поняття індексу клієнтоорієнтованості медичного закладу. В основу цього індексу покладемо інтегрований критерій якості медичних послуг, суть якого полягає у тому, що для вироблення якісної послуги необхідно, щоб вона надавалася компетентним працівником, у межах його робочого навантаження за необхідний час. Вважатимемо, що заклад є клієнтоорієнтованим, якщо всі послуги, які надаються його кадровим складом є якісними. Для визначення індексу клієнтоорієнтованості необхідно аналізувати співвідношення між нормативним і фактичним навантаженням працівників закладу.

Розглянемо заклад охорони здоров'я. Нехай задано множину працівників цього закладу $W = \{W_1, W_2, \dots, W_N\}$, для яких відомі значення нормативного (q_{iT}) та фактичного (r_{iT}) навантажень за вказаний період часу T ($i = \overline{1, N}$). З огляду на важливість і необхідність забезпечення якісними та своєчасними медичними послугами всіх споживачів медичних послуг, відповідно до принципу егалітаризму, індекс клієнтоорієнтованості закладу обчислюватимемо за правилом:

$$IC = \min \left\{ \min_{i=1, N} \left\{ \frac{q_{iT}}{r_{iT}} \cdot 100 \% \right\}, 100 \% \right\}. \quad (19)$$

Очевидно, що заклад буде максимально клієнтоорієнтованим, якщо $IC = 100\%$, тобто, якщо фактичне навантаження кожного працівника не перевищуватиме його нормативне навантаження, а, отже, у працівників буде достатньо обсягів кредитів часу для надання медичних послуг.

Оцінкою ефективності рішення буде величина, яка характеризує зміну індексу клієнтоорієнтованості й обчислюється так:

$$\Delta IC = IC' - IC, \quad (20)$$

де IC' – індекс клієнтоорієнтованості закладу охорони здоров'я після впровадження рішень, отриманих на основі застосування інформаційної технології.

Таким чином, для оцінювання ефективності розроблених моделей і методів, необхідно для заданого закладу охорони здоров'я виконати ситуаційний аналіз із визначення нормативного та



фактичного навантажень на працівників закладу, а також здійснити прогнозування потреб у медичних послугах на майбутні періоди часу. На основі отриманих результатів, застосовуючи модель пошуку оптимального розподілу послуг між працівниками, можна напрацювати варіанти управлінських рішень щодо зведення показника негативної клієнтоорієнтованості медичного закладу до нуля.

**ЕКСПЕРИМЕНТАЛЬНА
ВЕРИФІКАЦІЯ РОЗРОБЛЕНОЇ МОДЕЛІ**

Розглянемо ЗОЗ, у якому працює 13 працівників, які можуть надавати 7 різних послуг. Дані про розподіл послуг між працівниками наведено у табл. 1; табл. 2 містить дані про обсяги послуг.

ТАБЛИЦЯ 1
МАТРИЦЯ РОЗПОДІЛУ НАВАНТАЖЕННЯ
МІЖ ПРАЦІВНИКАМИ ЗАКЛАДУ

Працівник	Обсяги послуг, %						
	S_1	S_2	S_3	S_4	S_5	S_6	S_7
W_1			50		25		20
W_2			50	20	25		20
W_3	5			20	25		10
W_4	5				25		5
W_5		33					
W_6		33					
W_7				10		100	5
W_8		34					7
W_9							15
W_{10}				50			
W_{11}	40						
W_{12}	10						10
W_{13}	40						8

ТАБЛИЦЯ 2
ДАНІ ПРО ПОСЛУГИ

Послуга	Обсяг, кредити	Хронометраж, кредити
S_1	14937	10,3
S_2	14937	9,42
S_3	13714	10,36
S_4	2607,8	44
S_5	683	81
S_6	2733,1	11,2
S_7	5184	1

Результати обчислення фактичного і нормативного навантажень на працівників закладу [4, 5] занесено в табл. 3.

З даних табл. 3 випливає, що у працівників W_1 , W_2 , W_{10} фактичне навантаження значно перевищує нормативне. Це може призвести до перенавантаження працівників і зменшення

якості послуг, які вони надають. Результати порівняння навантаження на працівників показано на рис. 1.

ТАБЛИЦЯ 3
РЕЗУЛЬТАТИ ОБЧИСЛЕННЯ ФАКТИЧНОГО
ТА НОРМАТИВНОГО НАВАНТАЖЕНЬ НА ПРАЦІВНИКІВ ЗАКЛАДУ

Працівник	Нормативне навантаження, кредити	Фактичне навантаження, кредити
W_1	70467,55	85899,21
W_2	70467,55	108846,7
W_3	35233,78	44989,17
W_4	18167,69	21782,51
W_5	54304,57	46433,16
W_6	54304,57	46433,16
W_7	54304,57	42343,65
W_8	54304,57	48203,1
W_9	27146,89	777,6
W_{10}	31434,48	57368,67
W_{11}	62868,96	61540,44
W_{12}	15717,24	15903,51
W_{13}	62868,96	61955,16
Разом	611591,38	642476,04

На основі даних табл. 3 індекс клієнтоорієнтованості закладу: $IC \approx 54,8\%$.

Оскільки сумарне навантаження перевищує нормативне, то перерозподіл послуг між працівниками неможливий. У ході розв'язання задачі (14)–(18), шляхом зміни значень δp_{jT} та δr_{iT} , напрацьовано два варіанти рішень.

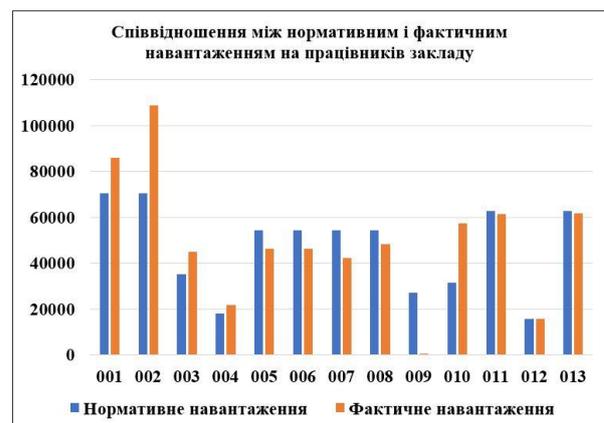


Рис. 1. Результати порівняння нормативного та фактичного навантаження на працівників закладу

Рішення 1: додати по одній штатній одиниці, ідентичній W_1 та W_7 . Варіант розподілу послуг між працівниками в такому випадку наведено у табл. 4.



ТАБЛИЦЯ 4

РОЗПОДІЛ ПОСЛУГ
МІЖ ПРАЦІВНИКАМИ ЗАКЛАДУ (РІШЕННЯ 1)

W _i	Послуги						
	S ₁	S ₂	S ₃	S ₄	S ₅	S ₆	S ₇
W ₁	1607		4301				
W _{1,1}	1607		4301				
W ₂			5112	4	87		
W ₃				116	372		
W ₄					224		
W ₅		4764					
W ₆		4707					
W ₇				886		1367	
W _{7,7}				887		1366	
W ₈		4085					5184
W ₉		1381					
W ₁₀				714			
W ₁₁	5103						
W ₁₂	525						
W ₁₃	6095						

Результати порівняння фактичного та нормативного навантаження за реалізації Рішення 1 зображено на рис. 2.

Індекс клієнтоорієнтованості ЗОЗ за реалізації Рішення 1 становитиме 100 %, проте у деяких працівників спостерігається значне вивільнення кредитів часу.



Рис. 2. Результати порівняння нормативного та фактичного навантаження на працівників закладу (Рішення 1)

Нехай у результаті послідовного розв'язання задач кластеризації, класифікації, ідентифікації та прогнозування обчислено майбутні потреби в медичних послугах, які наведено в табл. 5.

Для реалізації клієнтоорієнтованої оптимізації кадрового складу закладу на наступний рік, на основі результатів оцінювання фактичного і нормативного навантаження на працівників, матриці розподілу послуг між ними, а також резуль-

татів прогнозування потреб у медичних послугах розглянемо таку задачу цілочислового лінійного програмування, відповідно до (14)–(18):

ТАБЛИЦЯ 5

ПРОГНОЗНІ ЗНАЧЕННЯ ПОТРЕБ
У МЕДИЧНИХ ПОСЛУГАХ

Послуга	Обсяг, кредити	Хронометраж, кредити
S ₁	166839,4	10,3
S ₂	152585,12	9,42
S ₃	156317,284	10,36
S ₄	119269,3	44
S ₅	80757	81
S ₆	39335,52	11,2
S ₇	5184	1

$$F = 10,3 \cdot x_{1,1} + 10,3 \cdot x_{2,1} + 10,3 \cdot x_{3,1} + 10,3 \cdot x_{4,1} + 10,3 \cdot x_{10,1} + 10,3 \cdot x_{11,1} + 10,3 \cdot x_{12,1} + 10,3 \cdot x_{13,1} + 9,42 \cdot x_{5,2} + 9,42 \cdot x_{6,2} + 9,42 \cdot x_{7,2} + 9,42 \cdot x_{8,2} + 9,42 \cdot x_{9,2} + 10,359 \cdot x_{1,3} + 10,359 \cdot x_{2,3} + 44 \cdot x_{1,4} + 44 \cdot x_{2,4} + 44 \cdot x_{3,4} + 44 \cdot x_{4,4} + 44 \cdot x_{7,4} + 44 \cdot x_{10,4} + 44 \cdot x_{11,4} + 81 \cdot x_{1,5} + 81 \cdot x_{2,5} + 81 \cdot x_{3,5} + 81 \cdot x_{4,5} + 11,2 \cdot x_{7,6} + x_{1,7} + x_{2,7} + x_{3,7} + x_{4,7} + x_{5,7} + x_{6,7} + x_{7,7} + x_{8,7} + x_{9,7} + x_{12,7} + x_{13,7} \rightarrow \max,$$

$$x_{1,1} + x_{2,1} + x_{3,1} + x_{4,1} + x_{10,1} + x_{11,1} + x_{12,1} + x_{13,1} \geq 16198,$$

$$x_{5,2} + x_{6,2} + x_{7,2} + x_{8,2} + x_{9,2} \geq 16198,$$

$$x_{1,3} + x_{2,3} \geq 15090,$$

$$x_{1,4} + x_{2,4} + x_{3,4} + x_{4,4} + x_{7,4} + x_{10,4} \geq 2711,$$

$$x_{1,5} + x_{2,5} + x_{3,5} + x_{4,5} \geq 997,$$

$$x_{7,6} \geq 3512,$$

$$x_{1,7} + x_{2,7} + x_{3,7} + x_{4,7} + x_{5,7} + x_{6,7} + x_{7,7} + x_{8,7} + x_{9,7} + x_{12,7} + x_{13,7} \geq 432,$$

$$10,3 \cdot x_{1,1} + 10,359 \cdot x_{1,3} + 44 \cdot x_{1,4} + 81 \cdot x_{1,5} + x_{1,7} \leq 70467,55 + \delta r_1,$$

$$10,3 \cdot x_{2,1} + 10,359 \cdot x_{2,3} + 44 \cdot x_{2,4} + 81 \cdot x_{2,5} + x_{2,7} \leq 70467,55 + \delta r_2,$$

$$10,3 \cdot x_{3,1} + 44 \cdot x_{3,4} + 81 \cdot x_{3,5} + x_{3,7} \leq 35233,78 + \delta r_3,$$

$$10,3 \cdot x_{4,1} + 44 \cdot x_{4,4} + 81 \cdot x_{4,5} + x_{4,7} \leq 18167,69 + \delta r_4,$$

$$9,42 \cdot x_{5,2} + x_{5,7} \leq 54304,57 + \delta r_5,$$

$$9,42 \cdot x_{6,2} + x_{6,7} \leq 54304,57 + \delta r_6,$$

$$9,42 \cdot x_{7,2} + 44 \cdot x_{7,4} + 11,2 \cdot x_{7,6} + x_{7,7} \leq 54304,57 + \delta r_7,$$

$$9,42 \cdot x_{8,2} + x_{8,7} \leq 54304,57 + \delta r_8,$$

$$9,42 \cdot x_{9,2} + x_{9,7} \leq 27146,89 + \delta r_9,$$



$$10,3 \cdot x_{10,1} + 44 \cdot x_{10,4} \leq 31434,48 + \delta r_{10},$$

$$10,3 \cdot x_{11,1} + 44 \cdot x_{11,4} \leq 62868,96 + \delta r_{11},$$

$$10,3 \cdot x_{12,1} + x_{12,7} \leq 15717,24 + \delta r_{12},$$

$$10,3 \cdot x_{13,1} + x_{13,7} \leq 62868,96 + \delta r_{13},$$

$$x_{i,j} \in N \cup \{0\}, i = \overline{1,13}, j = \overline{1,7}.$$

У результаті розв'язання поставленої задачі було напрацьоване рішення, відповідно до якого, для підвищення індексу клієнтоорієнтованості закладу, необхідно залучити до роботи трьох додаткових працівників. Варіант розподілу навантаження між працівниками в такому випадку наведено у табл. 6.

Прогнозні значення фактичного та нормативного навантаження на працівників закладу наведено в табл. 7 і на рис. 3.

ТАБЛИЦЯ 6

**ПРОГНОЗНИЙ РОЗПОДІЛ ПОСЛУГ
МІЖ ПРАЦІВНИКАМИ ЗАКЛАДУ**

W _i	Послуги						
	S ₁	S ₂	S ₃	S ₄	S ₅	S ₆	S ₇
W ₁			6552				2592
W _{1,1}			6553				2592
W ₂			1985		616		
W ₃	1232			211	164		
W _{3,1}	1232			211	163		
W ₄					225		
W ₅		4765					
W ₆		4765					
W ₇				787		1756	
W _{7,1}				787		1756	
W ₈		4765					
W ₉		1903					
W ₁₀				714			
W ₁₁	6104						
W ₁₂	1526						
W ₁₃	6104						

ТАБЛИЦЯ 7

**РЕЗУЛЬТАТИ ОБЧИСЛЕННЯ ФАКТИЧНОГО
ТА НОРМАТИВНОГО НАВАНТАЖЕННЯ НА ПРАЦІВНИКІВ ЗАКЛАДУ**

Працівник	Нормативне навантаження, кредити	Фактичне навантаження, кредити
W ₁	70464,2	70467,55
W _{1,1}	70474,5	70467,55
W ₂	70458,6	70467,55
W ₃	35257,6	35233,78
W _{3,1}	35176,6	35233,78
W ₄	18225	18167,69

Закінчення табл. 7

Працівник	Нормативне навантаження, кредити	Фактичне навантаження, кредити
W ₅	44886,3	54304,57
W ₆	44886,3	54304,57
W ₇	54295,2	54304,57
W _{7,1}	54295,2	54304,57
W ₈	44886,3	54304,57
W ₉	17926,3	27146,89
W ₁₀	31416	31434,48
W ₁₁	62871,2	62868,96
W ₁₂	15717,8	15717,24
W ₁₃	62871,2	62868,96
Разом	734108,3	771597,28



Рис. 3. Результати порівняння прогнозного та нормативного навантаження на працівників закладу

Прогнозне значення індексу клієнтоорієнтованості за реалізації цього рішення буде $IC' \approx 99,7\%$, а $\Delta IC \approx 44,9\%$.

ВИСНОВКИ

Дослідження присвячено проблемі розроблення моделей і методів для реалізації клієнтоорієнтованої оптимізації кадрового складу закладів охорони здоров'я. Заклад охорони здоров'я можна розглядати як складну соціальну систему, основним структурним елементом якої є кадровий склад. Основними числовими характеристиками, якими можна оцінити кадровий потенціал ЗОЗ є фактичне та нормативне навантаження на працівників закладу. Відповідно до розроблених моделей і методів, навантаження обчислюється у кредитах – умовних одиницях часу з урахуванням індивідуальних ознак працівників закладу та особливостей надання окремих медичних послуг. Ключовим у формуванні кадрової політики в медичних закладах є етап прогнозування потреб у медичних послугах. Для його ефективної



реалізації доцільним є послідовне розв'язання задач кластеризації, класифікації, ідентифікації та прогнозування. На основі результатів оцінювання кадрового потенціалу ЗОЗ, та з урахуванням результатів прогнозу потреб у медичних послугах, можна розробити оптимальний розподіл навантаження між працівниками закладу. Для цього в дослідженні побудовано модель у вигляді задачі цілочислового лінійного програмування. Після розв'язання сформульованої задачі, з використанням розроблених продукційних правил, можливим є напрацювання множини управлінських рішень щодо клієнтоорієнтованої оптимізації кадрового складу ЗОЗ. Для оцінювання ефективності вироблених рішень уведено поняття індексу клієнтоорієнтованості закладу охорони здоров'я, в основу якого покладено інтегральний критерій якості медичної послуги. Пропонується обчислювати вказаний індекс, урахувавши принцип егалітаризму для забезпечення якості всіх медичних послуг, які надаються працівниками закладу.

У ході виконаної верифікації розроблених моделей продемонстровано спосіб вироблення варіантів рішень та оцінювання їхньої ефективності.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

- [1] В. І. Чепіженко. "Підхід до управління функціональним станом складних технічних систем на експлуатаційному інтервалі їх життєвого циклу," *Вісник Національного Авіаційного Університету*, 43(2), pp.53–57, 2010.
- [2] В. А. Грабовський, П. М. Клименко. "Системний підхід до управління закладами охорони здоров'я," *Вісник Національної академії державного управління при Президентові України*, (3), pp. 136–142, 2014.
- [3] O. Mulesa, V. Snytyuk, V. Nazarov. "Research of information-analytical aspects for optimization of the health care institutions," *Technology audit and production reserves*, 6(2 (50)), pp. 10–13, 2019.
- [4] O. Mulesa, F. Geche, V. Nazarov, M. Trombola, "Development of models and algorithms for estimating the potential of personnel at health care institutions," *Восточно-Европейский журнал передовых технологий*, (4 (2)), 52–59, 2019.
- [5] O. Mulesa, F. Geche, A. Batyuk, I. Myronyuk. "Using A Systematic Approach in the Process of the Assessment Problem Analysis of the Staff Capacity Within the Health Care Institution," In 2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT), Vol. 1, 2018, pp. 177–180.
- [6] Аналіз робочого навантаження, мотиваторів та стимулів праці медичних працівників, які надають віл-послуги в семи регіонах України. Проект USAID «Реформа ВІЛ-послуг у дії», Київ, 2017.
- [7] О. М. Крекотень. "Оцінка працезатрат лікаря загальної практики/сімейного лікаря на амбулаторному прийомі методом безпосередніх замірів часу через хронометраж," *Україна. Здоров'я нації*, № 3, pp. 61–67, 2014.
- [8] С. А. Батуркин, Е. Ю. Батуркина, В. А. Зименко, И. В. Сигинов. "Статистические алгоритмы кластеризации

данных в адаптивных обучающих системах," *Вестник РГ-ПТУ*, № 1 (31), pp.82–85, 2012.

[9] К. Фукунага К. *Введение в статистическую теорию распознавания образов*, М.: Наука. Главная редакция физико-математической литературы, 1979.

[10] А. Котов, Н. Красильников. *Кластеризация данных*, 2006. Режим доступа: yury.name/internet/02ia-seminar-note.pdf.

[11] H. Nguyen, C. Drebenstedt, X. N. Bui, D. T. Bui. "Prediction of blast-induced ground vibration in an open-pit mine by a novel hybrid model based on clustering and artificial neural network," *Natural Resources Research*, 29(2), 691–709, 2020.

[12] А. Г. Ивахненко. "Метод группового учета аргументов – конкурент метода стохастической аппроксимации," *Автоматика*, № 3, pp. 58–72, 1968.

[13] Ю. П. Зайченко, И. О. Заец, О. В. Камоцкий, О. В. Павлюк, "Исследование разных видов функций принадлежности параметров нечетких прогнозирующих моделей в нечетком методе группового учета аргументов," *Управляющие системы и машины*, 2, pp. 56–67, 2003.

[14] Ю. П. Зайченко. "Нечеткий метод группового учета аргументов при неопределенных входных данных," *Системні дослідження та інформаційні технології*, № 3, pp. 100–112, 2007.

[15] А. С. Мендель. "Метод аналогов в прогнозировании коротких временных рядов: экспертно-статистический подход," *Автомат. и телемех.*, № 4, pp. 143–152, 2004.

[16] Ю. П. Зайченко, Моамед Мухамед, Н. В. Шаповаленко. "Нечіткі нейронні мережі і генетичні алгоритми в задачах макроекономічного прогнозування," *Наукові вісті НТУУ "КПІ"*, № 4, pp. 20–30, 2002.

[17] C. Deb, F. Zhang, J. Yang, S. E. Lee, K. W. Shah. "A review on time series forecasting techniques for building energy consumption," *Renewable and Sustainable Energy Reviews*, 74, pp. 902–924, 2017.

[18] O. Mulesa, F. Geche, A. Batyuk, V. Buchok, V. "Development of combined information technology for time series prediction," In *Conference on Computer Science and Information Technologies*, 2017, pp. 361–373. Springer, Cham.

[19] О. Ю. Мулеса, В. Є. Снитюк. "Розробка еволюційного методу для прогнозування часових рядів," *Автоматизація технологічних і бізнес-процесів*. Том 12, вип. 3, pp. 4–9, 2020.

[20] О. Ю. Мулеса, В. Є. Снитюк, І. С. Миронюк. "Інформаційна технологія оптимізації кадрового потенціалу закладів охорони здоров'я," *Вісник Вінницького політехнічного інституту*. № 6, pp. 83–90, 2019.

Стаття надійшла до редколегії

14.09.2021



Customer-oriented optimization of healthcare facilities staff

The problem of developing information technology for customer-oriented optimization of healthcare facilities staff is considered. It is determined that the key tasks in the development of personnel decisions in the formation of personnel policy in medical institutions are assessing such current characteristics of staff as to regulatory and actual workload and the task of forecasting needs for medical services in future periods. To effectively perform such a forecast, it is proposed to consistently solve clustering, identification, and forecasting problems. A mathematical model of the multicriteria optimization problem for the formation of many variants of solutions for the formation of personnel policy in health care facilities is built. The model of search of optimum distribution of services between employees of establishment in the form of a problem of integer linear programming is executed. Production rules have been developed to analyze the results of solving the formulated problem. A model for developing variants of decisions on changes in the staff of a medical institution to ensure the completeness of medical services is proposed. To assess the effectiveness of the options made, the customer orientation index is used, which is calculated by the principles of egalitarianism. Experimental verification of the developed models and methods is performed.

Keywords: *decision-making, optimization, index of the customer orientation for the institution, customer-oriented optimization, human resources.*



Оксана Мулеца. Спеціальність 05.13.06 – інформаційні технології, канд. техн. наук, доцент кафедри кібернетики і прикладної математики ДВНЗ «Ужгородський національний університет», Ужгород, Україна. Наукові інтереси: інформаційні технології, прийняття рішень, нечіткий аналіз, інтелектуальний аналіз даних.

Oksana Mulesa. Candidate of Technical Sciences, specialty 05.13.06 – Information Technologies, Associate Professor, Associate Professor of Cybernetics and Applied Mathematics, Uzhhorod National University. Research interests include information technology, decision making, fuzzy analysis, data mining.



Віталій Снитюк. Декан факультету інформаційних технологій, д-р техн. наук за спеціальністю 05.13.06 – інформаційні технології, професор Київського національного університету імені Тараса Шевченка, Київ, Україна. Наукові інтереси: прийняття рішень в умовах невизначеності, обчислювальний інтелект, еволюційне моделювання.

Vitaliy Snytyuk. Dean of the Faculty of Information Technology, Doctor of Technical Sciences on the specialty 05.13.06 – Information Technology, Professor.

Research interests: decision making under uncertainty, computational intelligence, evolutionary modeling.



УДК 004.891

DOI: <https://doi.org/10.17721/AIT.2021.1.02>

О. В. Федусенко, orcid.org/0000-0002-5782-5922,
Н. Ю. Шкурпела,
І. М. Доманецька, orcid.org/0000-0002-8629-9933,
Київський національний університет імені Тараса Шевченка, Київ, Україна,
А. О. Федусенко, orcid.org/0000-0002-5782-5922,
ВЧ-К1410, Київ, Україна

Інтелектуальна система підтримки прийняття агротехнологічних рішень для посіву полів

У сучасному сільському господарстві України наявні проблеми, пов'язані з плануванням посівів. За допомогою запропонованої авторами інтелектуальної системи підтримки прийняття агротехнологічних рішень можливе спрощення процесу планування за рахунок використання концепції точного землеробства. Авторами проведено аналіз сучасних досліджень і публікацій, які пов'язані з концепцією точного землеробства і з проблемою впровадження сучасних інноваційних інформаційних систем у сільське господарство України. Проведено декомпозицію інтелектуальної системи та виокремлено шість основних підсистем, до кожної з яких розроблено функціональні вимоги. Проаналізовано сучасні методи моніторингу полів і визначено методи, що будуть використовуватися в інтелектуальній системі, одним з яких є метод *k*-середніх, який застосовуватиметься до кластеризації полів. На основі вже розроблених вимог авторами запропоновано загальну архітектуру системи, для графічного відображення якої використано нотацію TOGAF. Базуючись на запропонованій архітектурі, створено програмне забезпечення інтелектуальної системи. У результаті проведеного тестування програмного забезпечення інтелектуальної системи можна зробити висновок про її працездатність і готовність до впровадження. Спроектвана та розроблена система надає змогу здійснювати інтелектуальний аналіз історичних даних посівних, відображати результати у вигляді таблиць і графіків, виконувати планування посівів, агротехнологічних операцій і внесення добрив. Упровадження зазначеної системи дозволить підвищити якість прийняття управлінських рішень і продуктивність сільськогосподарської діяльності.

Ключові слова: інтелектуальна система, точне землеробство, кластерний аналіз, агротехнологічні рішення.

Для цитування (for citation): О. В. Федусенко, Н. Ю. Шкурпела, І. М. Доманецька, А. О. Федусенко. "Інтелектуальна система підтримки прийняття агротехнологічних рішень для посіву полів", *Сучасні інформаційні технології*, vol. 1, p. 15–22, 2021.

ВСТУП

Постановка проблеми. Сільське господарство є однією з провідних галузей економіки України, одним зі стратегічних напрямів розвитку якої є впровадження інноваційних технологій, зокрема адаптивних систем землекористування. Адаптивність у процесі прийняття виробничих та управлінських рішень забезпечує підвищення ефективності землеустрою та родючості ґрунту, продуктивності сільського господарства, а також зростання врожайності сільськогосподарських культур й істотне зниження виробничих витрат на сільськогосподарську продукцію.

Реалізація концепції точного землеробства спрямована на помітне підвищення ефективності аграрної галузі за рахунок організації своє-

часного та систематичного обліку, контролю й аналізу даних посівних площ за допомогою інтелектуальних систем підтримки прийняття рішень, які забезпечують їхню систематизацію та візуалізацію. Отримані результати дозволяють оптимізувати та здешевити процеси, пов'язані з плануванням посівів, розрахунком внесення норм добрив, прогнозуванням врожаю тощо. Тому актуальною науково-технічною задачею є дослідження та створення інтелектуальної системи підтримки прийняття агротехнологічних рішень для аналізу та візуалізації сільськогосподарських даних із використанням сучасних мов програмування та засобів візуалізації історичних даних щодо врожайності, агрохімічних досліджень ґрунту.

© Федусенко О. В., Шкурпела Н. Ю., Доманецька І. М., Федусенко А. О., 2021



Аналіз останніх досліджень і публікацій. Упровадження новітніх інформаційних технологій в аграрному секторі є важливою складовою оптимального управління та планування. Такий підхід дозволяє максимізувати прибутки та покращити організацію виробництва шляхом оптимізації використання господарських і природних ресурсів, електронного обліку посівів і зборів урожаїв, що дозволяє проводити точніше прогнозування майбутньої врожайності. Отже, унаслідок упровадження інновацій вдається отримати приріст виробництва сільськогосподарської продукції з одиниці площі без загрози довкіллю. За таких умов концепція традиційного землеробства вже не є ефективною.

Нині широкого розповсюдження набула концепція точного землеробства (precision agriculture). Упродовж кількох десятиліть вона успішно використовується та вдосконалюється країнами з високим рівнем розвитку аграрного виробництва, а саме: Канада, США, Нідерланди, Австралія тощо.

Наукова концепція точного землеробства базується на понятті про існування неоднорідностей у межах одного поля. Для оцінювання та визначення цих неоднорідностей використовують такі новітні технології: глобальні системи позиціонування (GPS, ГЛОНАСС), спеціальні датчики, аерофотознімки та супутникові знімки. Зібрані дані використовують для планування посіву, розрахунку внесення добрив і засобів захисту рослин згідно з нормами, більш точного прогнозування врожаю та фінансового планування. Ця концепція вимагає врахування локальних особливостей ґрунту (інформація про кількісну та якісну оцінку родючого шару), кліматичних умов [1].

Технологічні рішення є фундаментальною основою, яка дозволяє впроваджувати наукові дослідження та розроблення в аграрному секторі та в інших сферах розвитку промисловості. Тому питання планування й управління інноваційними засобами стали предметом наукових досліджень [2]. Проблеми, пов'язані з упровадженням цифрових технологій у виробничі та управлінські процеси сільського господарства, досліджували в наукових статтях такі українські й іноземні фахівці: М. І. Ромащенко, О. В. Шубравська, С. А. Володін, М. М. Кропивко, Tiziano Gomiero, E. C. Leonard, Keith Paustian, Ludwig Theuvsen та ін. [2–8].

Проте у зв'язку з динамічним розвитком інновацій в аграрній сфері можна стверджувати, що не всі проблеми пов'язані з управлінням інноваційним розвитком сільського господарства

знайшли широке відображення в наукових публікаціях. Багато аспектів упровадження точного землеробства, зокрема в сучасних умовах аграрного сектора України, досі обговорюються, оскільки спостерігається дефіцит усебічних досліджень. Це пов'язано з тим, що публікації, в яких досліджено переваги впровадження прецизійного землеробства, зосереджені лише на декількох аспектах, і, як правило, не містять практичної інформації.

ПОСТАНОВКА ЗАДАЧІ

Метою роботи є проектування та розроблення ефективної інтелектуальної системи підтримки прийняття агротехнологічних рішень для посіву полів для оптимального планування сівозмін і прийняття своєчасних управлінських рішень.

Для досягнення поставленої мети необхідно розв'язати такі задачі:

- дослідити основні методи моніторингу посіву полів та обрати ті, що будуть використовуватися в інтелектуальній системі;
- розробити архітектуру інтелектуальної системи підтримки прийняття агротехнологічних рішень для посіву полів;
- розробити програмне забезпечення інтелектуальної системи та провести його тестування.

Проведемо декомпозицію розроблюваної інтелектуальної системи, вона буде складатися з 6 підсистем. Для кожної з підсистем наведемо основні функціональні вимоги:

1. **Підсистема прогнозування та статистичного аналізу.** Система має забезпечувати функціональність прогнозування та статистичного аналізу з виконанням таких вимог:

- наявність функціонала кластерного аналізу;
- наявність функціонала статистичного аналізу;
- побудова прогнозування врожайності сільськогосподарських культур на основі історичних даних.

2. **Підсистема роботи з базою даних, а саме з основними довідниками.** Підсистема повинна забезпечувати можливість роботи з даними довідкового характеру з виконанням вимог:

- можливість додавання, видалення та редагування даних;
- можливість візуалізації даних у табличному вигляді;
- можливість налаштування користувачем правил відбору та сортування.

3. **Підсистема обліку врожайності сільськогосподарських культур.** Наведемо функціональні вимоги до системи щодо обліку врожайності сільськогосподарських культур:



- можливість ведення реєстру врожаїв сільськогосподарських культур за допомогою візуальних форм;
- можливість візуального відображення реєстру;
- забезпечення формування витягів із реєстру врожаю до звітів у розрізі років і видів сільськогосподарських культур;
- формування звітів про врожайність;
- можливість налаштування користувачем правил відбору, сортування у процесі формування аналітичних, звітних та інших таблиць;
- можливість імпорту/експорту даних.

4. Підсистема управління внесенням добрив. Ця підсистема має проводити моніторинг і надавати рекомендації щодо внесення добрив. Основними функціональними вимогами є:

- можливість ведення реєстру внесення добрив;
- розрахунок норм внесення добрив;
- можливість перегляду історії внесення добрив за полем і датою;
- формування рекомендацій щодо оптимізації розподілу добрив;
- можливість формування документа за результатами формування рекомендацій, що зберігаються у базі даних.

5. Підсистема обліку результатів фізико-хімічних аналізів ґрунту. Система повинна забезпечувати можливість обліку результатів фізико-хімічних аналізів ґрунту з виконанням таких вимог:

- ведення списку результатів фізико-хімічних аналізів ґрунту з прив'язкою до списку полів;
- можливість формування звіту фізико-хімічного стану ґрунту з вибором потрібного поля;
- визначення якості ґрунту шляхом розрахунку показника родючості.

6. Підсистема управління посівом полів.

Запишемо функціональні вимоги до системи щодо управління посівом полів:

- ведення списку полів із можливістю візуального відображення у вигляді таблиці;
- можливість ведення реєстру посівів із прив'язкою до списку полів;
- формування плану посівів;
- формування рекомендацій щодо оптимального засіву ріллі сільськогосподарськими культурами;
- можливість формування документа за результатами формування рекомендацій, які зберігаються у базі даних.
- можливість друку рекомендацій.

РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Аналіз методів моніторингу полів. Для оптимізації планування посіву полів спочатку бажано побудувати план розподілу ділянок і посіву полів. Одним із методів, який можна застосувати для розв'язання цієї задачі, є кластерний аналіз. Застосування методів кластерного аналізу даних дозволяє пришвидшити виконання процесу інтеграції посівних площ з аналогічними властивостями. Тобто, на основі даних фізико-хімічних властивостей ґрунту або врожайності сільськогосподарських культур, вдається виконати розподіл полів на групи за критерієм найбільшої подібності. Усі методи кластерного аналізу можна поділити на ієрархічні та неієрархічні. Неієрархічні методи виявляють вищу стійкість щодо шумів і викидів, некоректного вибору метрики, введення незначущих змінних у набір, що бере участь у кластеризації [9]. Основним недоліком цієї групи методів є те, що необхідно заздалегідь визначити кількість кластерів. Ієрархічні методи не потребують визначення точної кількості кластерів, тому що вони будують повне дерево вкладених кластерів. Проте застосування ієрархічних методів для великих наборів даних досить складне. З урахуванням проведеного порівняльного аналізу авторами прийнято рішення про використання неієрархічних методів кластеризації, а саме методу *k*-середніх [10], який є найпоширенішим серед групи неієрархічних методів. Указаний метод доцільно застосовувати для кластеризації полів, оскільки його застосування вимагає гіпотезу про кількість кластерів. Для побудови плану розподілу полів попередньо визначена кількість кластерів є важливим чинником, тому що надає можливість установити обмеження на кількість груп, яка не може бути невизначеною або мати дуже великий розмір.

Одним із найважливіших показників, які визначають рівень ефективності ведення сільськогосподарського виробництва, є врожайність сільськогосподарських культур. Цей показник використовується як вхідна інформація для побудови планів, прогнозів і прийняття управлінських рішень, а також є одним з основних результуючих показників сільськогосподарського виробництва. Отримання достовірного прогнозу врожаю забезпечує оптимізацію об'ємів і структури резервних фондів і запасів, поліпшення ведення політики зовнішньої торгівлі, ефективне маневрування структурою та розміщенням виробництва.

Найуспішніші прогнози реалізуються, коли побудована адекватна математична модель об'єкта.



Для моделювання динаміки врожайності найбільше поширення здобули лінійні стохастичні моделі, основою яких є аналіз часових рядів. Раніше проведені дослідження показали, що часовим рядам урожайності притаманні короткі цикли тривалістю 4 роки і середні цикли тривалістю 16–20 років [11]. Короткий цикл швидше за все викликаний циклічністю погоднокліматичних факторів, середній може бути пояснений у межах моделі «врожайність–родючість», яка є моделлю типу «хижак–жертва» [11].

Автори пропонують використовувати два методи для прогнозування врожайності, а саме: метод багатofакторної лінійної регресії, використання якого докладніше описано у [12], та алгоритм ARIMA.

Алгоритм ARIMA – це реалізація процесу обчислення авторегресивного інтегрованого ковзного середнього, описаного в роботі Бокса – Дженкінса. Методологія ARIMA дозволяє визначити залежності в результатах спостережень у формі послідовних часових рядів, за умови, що модель включає опис випадкових стрибків [12].

Оптимальне внесення добрив для живлення рослин є ще одним із найважливіших чинників системи землеробства. Внесення необхідних норм добрив забезпечує отримання високих урожаїв хорошої якості у разі підвищення або збереження досягнутого рівня родючості ґрунту, і не завдаючи шкоди навколишньому природному середовищу. Окрім того, оптимальний розподіл добрив сприяє економічній ефективності сільськогосподарського виробництва. Чинниками, що формують родючість ґрунтів є сукупність фізичних, хімічних і біологічних властивостей ґрунтів. Якраз хімічні властивості родючості ґрунтів ураховують під час розрахунку й оптимізації внесення добрив у інтелектуальній системі, а саме: вміст у ґрунті азоту, фосфору і калію, а також інших поживних елементів.

Норми внесення добрив визначають такими методами:

- 1) використання результатів польових дослідів із добривами;
- 2) балансово-розрахунковий;
- 3) комплексний;
- 4) економіко-математичний.

Найбільшого поширення набули балансово-розрахункові методи, зокрема розрахунок норм добрив за запасами поживних речовин у ґрунті на запланований урожай і розрахунок норм добрив на приріст урожаю [13–16]. Зазначені методи і будуть використані у розроблюваній системі підтримки прийняття рішень для розрахунку оптимальних норм унесення добрив під основні сільськогосподарські культури.

Розроблення архітектури інтелектуальної системи. Перед побудовою архітектури опишемо основні структурні компоненти системи підтримки прийняття рішень агротехнологічних рішень для посіву полів.

Інтерфейс користувача – дозволяє наочно продемонструвати різні довідники, що зберігаються в системі, переглянути звітні матеріали, а також переглянути інформацію, отриману за допомогою інформаційно-пошукового модуля, відобразити дані, отримані в результаті інтелектуального аналізу даних; забезпечує зручність введення даних користувачем у базу даних.

Підсистема управління даними – зберігає інформацію про посівні ріллі, урожайність, результати фізико-хімічного аналізу ґрунтів, дані стосовно норм внесення добрив і довідкові значення та показники, необхідні для розрахункових операцій. Окрім того, передбачена можливість збереження рекомендацій щодо посіву полів. Уся інформація для кращого візуального сприйняття передається у програмну оболонку. Ця підсистема складається з БД певної структури та СУБД, яка забезпечує швидке виконання аналітичних запитів.

У зазначеній підсистемі використовують два типи джерел даних:

- довідкові дані, які містять значення коефіцієнтів та інших стандартних показників, необхідних для проведення розрахункових операцій. Джерелом даних є нормативні документи у галузях ґрунтознавства, агрохімії та охорони ґрунтів.
- внутрішні дані, які вводяться безпосередньо користувачем, а саме: показники врожайності, результати фізико-хімічного аналізу ґрунту тощо. Тобто це дані, які є результатом сільськогосподарської діяльності. Джерелом даних є нормативні документи у галузях ґрунтознавства, агрохімії та охорони ґрунтів і дані конкретного сільськогосподарського підприємства.

Інформаційно-аналітична підсистема – дозволяє проводити розгорнутий аналіз даних і складається з таких компонентів: підсистема інтелектуального аналізу даних, підсистема управління внесенням добрив і підсистема управління посівними площами.

Підсистема інтелектуального аналізу – реалізує методи й алгоритми здобуття даних Data Mining, а саме: метод k -середніх, який належить до методів неієрархічного кластерного аналізу, та лінійну багатofакторну регресію. Інструменти інтелектуального аналізу дозволяють виявити закономірності в діяльності сільськогосподарського виробництва, використовувані надалі для обґрунтування стратегічних або тактичних рішень. Ця підсистема складається з двох модулів: модуль



прогнозування врожайності сільськогосподарських культур і модуль кластерного аналізу полів із метою оптимального планування їхнього посіву.

Модуль прогнозування врожайності сільськогосподарських культур – дозволяє прогнозувати врожай у майбутньому на основі даних, які зберігаються в БД.

Модуль кластерного аналізу полів з метою оптимального планування їх посіву – дозволяє виконати перерозподіл посівних площ задля оптимального планування їх посіву, враховуючи дані врожайності, результати фізико-хімічних аналізів ґрунтів.

Підсистема управління внесенням добрив – забезпечує введення та обробку даних стосовно добрив.

Підсистема управління посівними площами – забезпечує процес обробки даних щодо врожайності, результатів фізико-хімічних аналізів ґрунтів, формування звітної документації та рекомендацій щодо посіву полів. Зазначена підсистема складається з наступних підсистем:

- підсистема обліку врожайності – містить модуль ведення реєстру врожайності та модуль формування звітів по врожайності;
- підсистема обліку результатів фізико-хімічних аналізів ґрунту;
- підсистема управління посівом полів.

Графічне представлення загальної архітектури розроблюваної системи підтримки прийняття рішення представлено у нотатції архітектурного фреймворку TOGAF наведено на рис. 1.

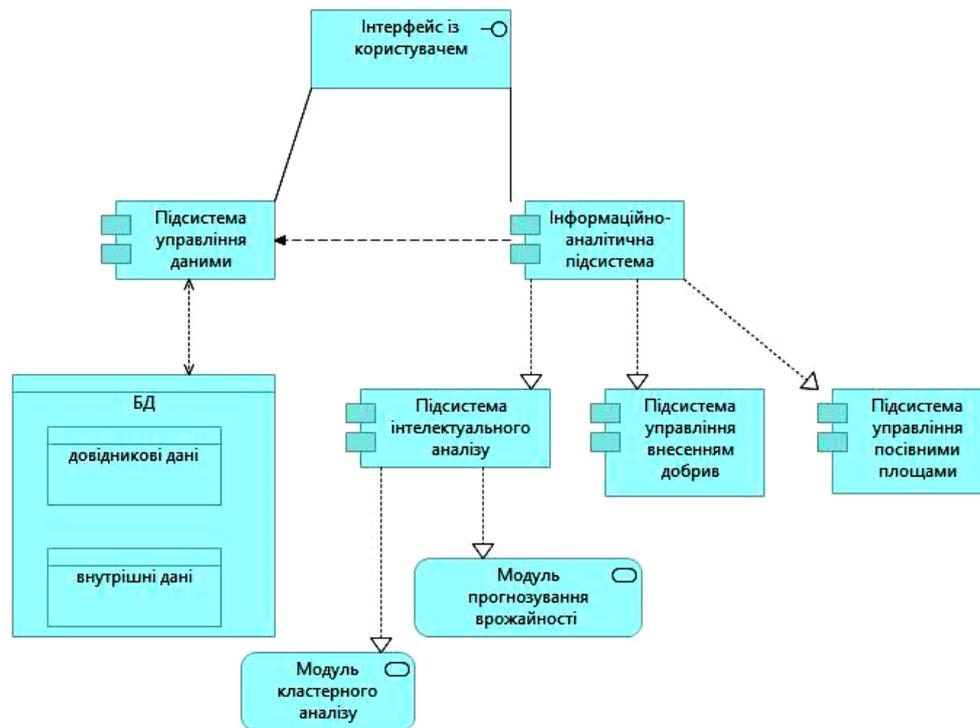


Рис. 1. Загальна архітектура інтелектуальної системи

Розроблення програмного забезпечення інтелектуальної системи. Програмне забезпечення інтелектуальної системи підтримки прийняття агротехнологічних рішень складається з набору програмних модулів, які за змістом є ідентичними до архітектурних компонентів, представлених на рис. 1.

Модулі мають трирівневу архітектуру, яка передбачає наявність презентаційного рівня, тобто клієнтської програми, рівня управління бізнес-логікою та рівня доступу до даних за рахунок системи управління базою даних. Тому задля зручного управління об'єктами у процесі розроблення сис-

теми, відповідно до заданої моделі архітектури, було виділено чотири простори імен (рис. 2).

Розглянемо `AgricultureDataSetTableAdapters` та `Precision_agriculture` детальніше. `AgricultureDataSetTableAdapters` містить об'єкти `TableAdapters`, які забезпечують зв'язок між модулями підсистем і базою даних.

Простір імен `Precision_agriculture` реалізує одразу два рівні архітектури: оброблення даних і взаємодію з користувачем. Взаємодію з користувачем реалізовано за допомогою графічного інтерфейсу `Windows Forms`.

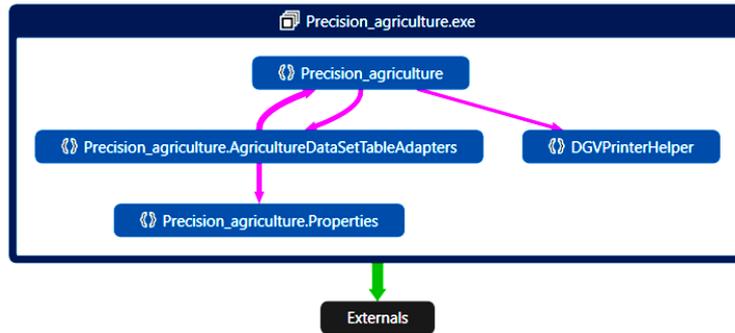


Рис. 2. Схема взаємодії між просторами імен інтелектуальної системи

Тестування програмного забезпечення інтелектуальної системи. Розглянемо приклад роботи з програмним забезпеченням інтелектуальної системи.

Спочатку проведемо аналіз роботи підсистеми управління внесенням добрив, що дозволяє виконувати операції з маніпулювання даними в реєстрі добрив, а саме: додавання нових, редагування існуючих, видалення. Передбачає візуальне їхнє відображення у табличному вигляді та можливість налаштування користувачем правил відбору й сортування. Надає можливість розрахунків із визначення норм поживних речовин, необхідних для внесення з метою отримання запланованого врожаю, та відповідних доз мінеральних добрив, а також визначити насиченість ґрунту органічними добривами та записати отримані дані до плану внесення добрив. Усі обчислення виконують за допомогою модуля «Розрахунок річних доз добрив».

Показники розраховують для обраних зі списку номера поля та року врожаю. За потреби є можливість внесення даних про органічні добрива, доза внесення яких впливає на визначення норм внесення мінеральних добрив.

Розглянемо результат розрахунку описаних вище показників за допомогою системи (рис. 3) та порівняємо їх зі значеннями, розрахованими за допомогою програми MS Excel (рис. 4). Для тестування використовуватимемо реальні сільськогосподарські дані, які наведено в [17].

Отримані результати є однаковими для обох випадків, що свідчить про правильну та коректну роботу інтелектуальної системи.

Проведемо тестування роботи модуля кластерного аналізу. Інструменти модуля надають можливість проводити аналіз для двох видів даних: агрохімічні показники ґрунтів і врожайність сільськогосподарських культур. Початкові дані та результати кластерного аналізу відображаються на відповідних графіках.

Проведемо тестування модуля кластеризації полів за агрохімічними характеристиками ґрунтів. За результатами аналізу дані поділено на

три кластери. На осі X відображаються дані вмісту азоту, а на осі Y – фосфору. Розріз даних по кластерах можна переглянути для кожного окремого показника, вибравши на вкладці СТАТИСТИКА для іншого показника значення 0 (рис. 5).

Рис. 3. Результати розрахунку системи

Попередник:	Горox	0,7	0,8	0,85
Культура:	Озима пшениця			
Мінеральне добриво:	Аміачна селітра			
Необхідно внести поживних речовин мінеральних добрив з урахуванням коефіцієнта їх використання, кг/га		248,00	-13,04	-34,52
Вміст поживних речовин у мінеральних добривах, %		0,35	1,00	1,00
Необхідно внести мінеральних добрив для одержання запланованого врожаю, ц/га		496	-10,4348	-29,3425

Рис. 4. Результати розрахунку системи за допомогою програми MS Excel

Кластер	1	2
Кластер: 0		
1	1,4	8
8	1,3	7
13	1,3	10
14	1,5	7
Кластер: 1		
2	1,5	7
7	1,5	7
9	1,2	8

Рис. 5. Результати роботи модуля кластеризації



Проведемо порівняльний аналіз результатів прогнозування врожайності сільськогосподарських культур, отриманих за допомогою інтелектуальної системи, та реальної врожайності цих культур, відповідно до даних, що наведені у [18] на прикладі Дніпропетровської області. Результати аналізу наведено у таблиці.

ТАБЛИЦЯ
РЕЗУЛЬТАТИ ПРОГНОЗУВАННЯ ВРОЖАЙНОСТІ

Назва культури	Прогнозоване інтелектуальною системою значення врожайності 2021 р., ц/га	Фактичне значення 2021 р., ц/га	Різниця, ц/га	Похибка, %
Кукурудза	43,06	45,9	2,84	-5,5
Пшениця	43,24	44,22	0,98	-2,3
Ячмінь	30,28	33,79	3,51	-11,6
Горох	22,38	24,01	1,63	-7,3

Як можна побачити з таблиці, результати прогнозу отримано з використанням інтелектуальної системи, майже відповідають фактичним результатам урожайності відповідних сільськогосподарських культур, наявна різниця є невеликою і складає менше, ніж 4 ц/га, а точність прогнозу складає приблизно 90 %, що є досить високим показником.

За результатами тестування визначено високу якість функціональних характеристик та атрибутів системи. Система проста у використанні, має зручний користувацький інтерфейс, реалізовує багато функціональних можливостей.

ВИСНОВКИ

Упровадження інноваційних інформаційних технологій у сільське господарство України є актуальною проблемою, для вирішення якої можуть бути використані адаптивні системи землекористування. Розроблена авторами інтелектуальна система підтримки прийняття агротехнологічних рішень реалізує концепцію точного землеробства та призначена для використання фермерськими господарствами для планування та моніторингу посівів полів.

Упровадження зазначеної системи дозволить підвищити якість прийняття управлінських рішень і продуктивність сільськогосподарської діяльності.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

[1] Федірець О. В. «Управління інноваціями при впровадженні технологій точного землеробства в Україні», *Наукові праці Полтавської державної аграрної академії*, т. 3, № 2 (7), с.302–308, 2013

[1] Зеліско Н. Мельник В. «Розвиток інноваційного потенціалу аграрного сектору економіки України». *Вісник Львівського національного аграрного університету. Сер. Економіка АПК* 2018. Вип. 25. С. 40–43.

[2] Tiziano Gomiero. (2019) Soil and crop management to save food and enhance food security. [Online]. Available: <https://doi.org/10.1016/B978-0-12-815357-4.00002-X>

[3] E.C.Leonard. (2015) Precision Agriculture. Reference Module in Food Science [Online]. Available: <https://doi.org/10.1016/B978-0-08-100596-5.00203-1>

[4] Diego de la Rosa, Francisco Mayol, Elvira Díaz-Pereira, Miguel Fernandez, Diego de la Rosa. (2004) A land evaluation decision support system (MicroLEIS DSS) for agricultural soil protection: With special reference to the Mediterranean region. [Online]. Available: <https://doi.org/10.1016/j.envsoft.2003.10.006>

[5] D. de la Rosa, F. Mayol, E. Diaz-Pereira, M. Fernandez and D. de la Rosa Jr. (2003) A land evaluation decision support system (MicroLEIS DSS) for agricultural soil protection. 2003. [Online]. Available: <https://doi.org/10.1016/j.envsoft.2003.10.006>

[6] (2018) Requirement types. QAInfo [Online]. Available: <https://www.quality-assurance-group.com/requirement-types/>

[7] Чубукова І.А., *Data Mining*, М.: Інтернет-Університет інформаційних технологій, БІНОМ. Лабораторія знань, 2006.

[8] І. І. Глаголева, А. Ю. Берко «Застосування кластерного аналізу для опрацювання даних земельного кадастру». *Вісник Львівського національного університету «Львівська політехніка». Інформаційні системи та мережі*. 2014. Вип.743. С. 420–429

[9] Nathan Landman, Hannah Pang, Christopher Williams, Eli Ross. (2017) k-Means Clustering. [Online]. Available: <https://brilliant.org/wiki/k-means-clustering/>

[10] Вітлінський В. В., Грицюк П. М. «Дослідження динаміки урожайності озимої пшениці для областей України». *Моделювання та інформаційні системи в економіці: Зб. наукових праць*. 2007. Вип. 76. С. 275–295.

[11] Gorban A.N., Zinovyev A.Y. «Principal Graphs and Manifolds», *Ch. 2 in: Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*, IGI Global. Hershey. PA. USA., p.28–59 2009.

[12] Грицюк П.М. «Прогнозування врожайності зернових культур: особливості і методика», *Вчені записки зб. наук. пр. ДВНЗ «Київський національний економічний університет імені Вадима Гетьмана»*, №11, с. 294–300, 2009

[13] Якимчук В. Г., Жолобак Г. М., Порушкевич А. Ю., Сахацький О. І. «Використання космічних і метеорологічних даних для оцінки врожайності озимої пшениці», *Космічна наука і технологія*, Т. 17, № 5, с. 64–67, 2011.

[14] А.С. Заришняк, *Оптимізація удобрення та родючості ґрунту в сівозмінах*, Київ: Аграрна наука, 2015.

[15] С.А. Балюк, Б.С. Носко, В.В. Шимель, Л.В. Стеревська, Г.Ф.Момот «Оптимізація живлення рослин у системі факторів ефективної родючості ґрунтів», *Вісник аграрної науки*, Т.7, №3, с.12–19, 2019

[16] (2021). Державна служба статистики України [Online]. Available: <http://www.ukrstat.gov.ua/>

[17] (2021). Врожай онлайн [Online]. Available: <https://latifundist.com/urozhaj-online-2021#>

Стаття надійшла до редколегії

06.10.2021



Intelligent support system for agro-technological decision-making for field sowing

The are crop planning problems exist in a modern agriculture of Ukraine. With the help of the intelligent support system for agro-technological decisions proposed by the authors, it is possible to simplify the planning process by using the concept of precision farming. Modern fields monitoring methods were analyzed and methods that will be used in the intelligent system are identified. The k-means method is one of them and will be applied to field clustering. The authors analyzed modern research and publications related to the concept of precision farming and the problem of implementing modern innovative information systems in agriculture of Ukraine. The decomposition of the intelligent system was carried out. Six main subsystems were identified, functional requirements were developed for each of them. Modern methods of fields monitoring are analyzed and methods that will be used in the intelligent system are identified, one of which is the k-means method, which will be applied to field clustering. Based on the already developed requirements, the authors have developed the general architecture of the system. The notation TOGAF was applied for the graphical display of the architecture. Based on the proposed architecture, intelligent system software was created. As a result of testing the soft-ware of the intelligent system, it is possible to draw a conclusion about its efficiency and readiness for implementation. The designed and developed system allows to carry out intellectual analysis of historical data of crops, to display results in the form of tables and graphs, to carry out planning of crops, agrotechnological operations and fertilizer application. The introduction of this system will improve the quality of management decisions and productivity of agricultural activities.

Keywords: intelligent system, precision farming, cluster analysis, agro-technological solutions.



Олена Федусенко. Працює доцентом кафедри інтелектуальних технологій Київського національного університету імені Тараса Шевченка, Київ, Україна, канд. техн. наук у галузі інформаційних технологій з 2006 р. Є автором понад 60 наукових праць. Наукові інтереси: системи адаптивного навчання, генетичні алгоритми, методи Data Mining.

Olena Fedusenko. Works as an associate professor of the Department of Intellectual Technologies of Taras Shevchenko National University of Kyiv, Kyiv, Ukraine. She got degree of candidate of technical sciences in the field of information technologies in 2006. He is the author of more than 60 scientific works. Research interests include adaptive learning systems, genetic algorithms and their application, Data Mining methods.



Наталія Шкурпела. Отримала освітній ступінь «бакалавр» та освітню кваліфікацію «Бакалавр комп'ютерних наук та інформаційних технологій» у Київському національному університеті імені Тараса Шевченка. Київ, Україна.

Natalia Shkurpela. She graduated with a bachelor's degree in Computer Science from Taras Shevchenko National University of Kyiv, Kyiv, Ukraine.



Ірина Доманецька. Працює на посаді доцента кафедри інтелектуальних технологій Київського національного університету імені Тараса Шевченка, канд. техн. наук. Київ, Україна. Сфера наукових досліджень – системотехнічні дослідження в галузі ІТ, нейромережні технології та їхнє застосування, адаптивні системи навчання.

Iryna Domanetska. Candidate of Technical Sciences, Associate Professor, works as an associate professor of the Department of Intellectual Technologies of Taras Shevchenko National University of Kyiv, Kyiv, Ukraine. Areas of research are system-technical research in the field of IT, neural network technologies and their application, adaptive learning systems.



Анатолій Федусенко. Військовослужбовець. Закінчив Київський національний університет будівництва та архітектури у 2000 р., отримавши кваліфікацію інженер з автоматизації, канд. техн. наук у галузі інформаційних технологій з 2016 р. Є автором багатьох наукових праць. Наукові інтереси: генетичні алгоритми та їхнє прикладне застосування, методи Data Mining і візуалізації даних, адаптивні системи навчання.

Anatoliy Fedusenko. Serviceman. He graduated from Kyiv National University of Construction and Architecture in 2000 with a degree in automation engineering. He received the degree of Candidate of Technical Sciences in the field of information technology in 2016. He is the author of many scientific works. Research interests include genetic algorithms and their application, Data Mining and data visualization methods, adaptive learning systems



ШТУЧНИЙ ТА ОБЧИСЛЮВАЛЬНИЙ ІНТЕЛЕКТИ



УДК 004.89

DOI: <https://doi.org/10.17721/AIT.2021.03>

Maryna Antonevych, orcid.org/0000-0003-3640-7630,
Anna Didyk,
Nataliia Tmienova, orcid.org/0000-0003-1088-9547,
Vitaliy Snytyuk, orcid.org/0000-0002-9954-8767,
Taras Shevchenko National University of Kyiv, Ukraine

Choosing the best parameters for method of deformed stars in n -dimensional space

This paper is devoted to the problem of optimization of a function in n -dimensional space, which, in general case, is polyextreme and undifferentiated. The new method of deformed stars in n -dimensional space was proposed. It is built on the ideas and principles of the evolutionary paradigm. Method of deformed stars is based on the assumption of using potential solutions groups. There by it allows to increase the rate of the accuracy and the convergence of the achieved result. Populations of potential solutions are used to optimize the multivariable function. In contrast to the classical method of deformed stars, we obtained a method that solves problems in n -dimensional space, where the population of solutions consists of 3-, 4-, and 5-point groups. The advantages of the developed method over genetic algorithm, differential evolution and evolutionary strategy as the most typical evolutionary algorithms are shown. Also, experiments were performed to investigate the best configuration of method of deformed stars parameters.

Keywords: enterprise, technology, optimization, method, experiment.

For citation: M. Antonevych, A. Didyk, N. Tmienova, V. Snytyuk, "Choosing the best parameters for method of deformed stars in n -dimensional space" *Advanced Information Technology*, vol. 1, p. 24–28, 2021.

INTRODUCTION

The large number of modern practical problems belong to the class of constraint satisfaction problems (CSPs). Such algorithms, as stochastic search, combinatorial optimization methods, and evolutionary algorithms are used to solve these tasks.

In artificial intelligence, the evolutionary algorithm (or EA) is a subset of evolutionary computations [1]. It is general population-based metaheuristic optimization algorithm. All evolutionary algorithms have the basic provisions in the theory of biological evolution - the processes of selection, mutation and reproduction. The environment determines the behavior of elements in population. Such a population evolves in accordance with the selection rules in accordance with the objective function set by the domain.

Thus, each individual in the population is assigned a value for its suitability in the domain. Only the most suitable species produce offspring. Various types of mutation allow agents to change and adapt to the domain. Thus, such algorithms are classified as adaptive search engines.

PRACTICAL USE

As it is known, evolutionary algorithms are used for problems such as functional optimization and can be easily described with mathematical formulas. EA are used in combinatorial optimization, particularly in solving classical NP-complete problems such as the traveling salesman problem (TSP) [2], number splitting, knapsack packing problem [3], maximally independent set, and graph sketching [4].

Also, other non-classical problems for which EA are used include transportation problems, route calculation, scheduling, planning, location, etc. Evolutionary algorithms are also used to optimize structures and electronic circuits in economics and medicine. In addition, evolutionary algorithms can be used in music and they are being actively explored in Austria, essentially when trying to simulate playing musical instruments by popular people from different years [5].

REVIEW OF EVOLUTIONARY METHODS FOR FUNCTION OPTIMIZATION

As it was said earlier, evolutionary algorithms can be used to solve optimization problems. The

© Antonevych M., Didyk A., Tmienova N., Snytyuk V., 2021



most famous algorithms are genetic algorithm (GA), evolutionary strategy (ES), evolutionary programming (EP), genetic programming (GP), etc.

Evolutionary programming was introduced by Lawrence J. Vogel [6]. Like ES and GA, EP has a practical optimization technique when other techniques such as gradient descent (GD) or direct analytic discovery (DAD) are not plausible.

Combinatorial optimization and optimization of functions with real values, in which the optimization surface or fitness landscape is "hard to pass", with many locally optimal solutions, are well suited for evolutionary programming.

Classical EPs have been developed to evolve finite-state automata in such a way that they were capable of solving time series prediction tasks and more generally evolving learning machines [7]. Modern EPs have been applied to solve continuous optimization problems using real-valued representations [7]. For example, to find the shortest path in a TSP, each solution would be a path. The length of the path (as example, from location a to location b) can be expressed as a number, which in turn would serve as the fitness criteria for the solution [8].

John H. Holland created the genetic algorithm [9]. A GA is a method for solving problems of both constrained and unconstrained optimization, based on the process of natural selection, which mimics biological evolution. GA can be applied to solve problems that are not well suited for standard optimization algorithms, including problems in which the objective function is non-differentiable, stochastic, discontinuous or highly nonlinear.

Ingo Rechenberg and Hans-Paul Schwefel developed an evolutionary strategy approach [10-11]. Evolutionary strategy is a stochastic global optimization algorithm. Also, ESs can be applied in all fields of optimization including continuous, discrete, combinatorial search spaces without and with constraints as well as mixed search spaces. The ESs can also be applied to a set of objective functions in context of multi-objective optimization.

In the early nineties, a fourth stream appeared - genetic programming. Genetic programming has been successfully used as a tool for automatic programming, automatic problem solving and machine learning. This algorithm is especially useful in areas where the exact shape of a solution is not known in advance or an approximate solution is acceptable (perhaps because it is very difficult to find an exact solution).

Some of the GP applications include symbolic regression, curve fitting, data modeling, classification, feature selection, etc. John R. Koza mentions 76 cases where genetic programming could produce

results that compete with results created by humans (the so-called human-competitive results).

Since the 1990s, evolutionary computing has largely become associated with the idea of swarm intelligence [12], and nature-inspired algorithms have become an increasingly significant part of this trend. Swarm intelligence (SI) is concerned with the design of intelligent multi-agent systems by taking inspiration from the collective behavior of social insects as well as from other animal societies [13].

Swarm intelligence methods have been very successful in the area of optimization, which is of great importance for industry and science [13]. Optimization problems are of high importance both for the industrial world as well as for the scientific world.

Examples of practical optimization problems: train scheduling, timetabling, shape optimization, problems from computational biology and telecommunication network design. The research community has simplified many of these problems in order to obtain scientific test cases such as the well-known TSP [13].

PROBLEM FORMULATION

Let's consider the problem of finding the minimum value of the multivariable function:

$$\begin{aligned} f(x_1, x_2, \dots, x_n) &\rightarrow \min, \\ \bar{x} &= (x_1, x_2, \dots, x_n), \\ \bar{x} &\in D \subset R^n, \end{aligned} \quad (1)$$

where n is the dimension of the search space, D is some area.

The problem (1) is equivalent to the search problem:

$$\begin{aligned} \text{Arg min } f(\bar{x}), \\ x \in D. \end{aligned} \quad (2)$$

In addition, let some unknown point $x^* = (x_1^*, x_2^*, \dots, x_n^*) \in D$ will be its solution.

Since the problem has an optimization character, the No Free Lunch Theorem (NFL) is applicable to it [14]. The meaning of this theorem is that there is no single method that can find the best solution to any optimization problem.

Thus, the use of evolutionary algorithms does not require strict target functions constraints and does not guarantee the finding of a global optimum. That is why both the enlargement of optimization methods and the research of their effectiveness continue.

Note that the function f can be undifferentiated, polyextreme, besides it can be given tabularly or algorithmically. Therefore, in such cases, the methods of classical optimization are ineffective. Hypothesis is that there is a motion direction of the potential solution of the problems (1), (2), which will

allow with high probability and quickly finding the optimal solution of the problem.

Moreover, there is an area in neighborhood of potential solution to be investigated in more detail because it has solutions that are better than the parent's potential solution [15].

We propose to use the idea that underlies all evolutionary algorithms, namely the population approach, which consists of solving problems (1), (2) by evolution the population of potential solutions [16].

The key elements of the solution search for the problems (1), (2) are special transformations like compressions, and rotations of figures formed by the connection of points, which are the potential solutions.

In addition, these groups of points are called deformed stars. For this reason, it is necessary to use a larger number of points grouped in a certain way and to make a directional movement towards finding a potential best solution with a detailed study of the surrounding area.

That is an idea underlying the method of deformed stars [17]. In this paper, the method is further developed by extending it into groups of 3, 4, and 5 potential solutions (3-, 4- and 5-point stars) [18].

THE PROPOSED ALGORITHM

Step 1. We initialize the parameters of the algorithm, let $t = 0$.

Step 2. Generate m potential solutions in domain D (population P_t).

Step 3. Form w figures (triangles, rectangles or pentagons) $F_i, i = \overline{1, w}$.

Step 4. For each F_i find the vertex in which the function f takes the best value and consider it the best vertex.

Step 5. For each figure F_i find a compressed figure T_i in which the best point is transferred along the line connecting the center of the figure and the best point, and all the others are transferred to it (Figure 1).

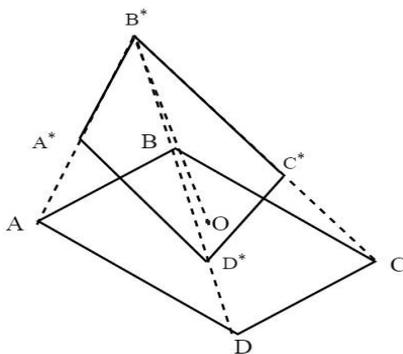


Fig. 1. Figure transformations

Step 6. For each figure F_i find a compressed figure U_i , in which all points are compressed to the best vertex.

Step 7. For each figure F_i find the figure Q_i , obtained by rotating F_i around the best vertex (Figure 2).

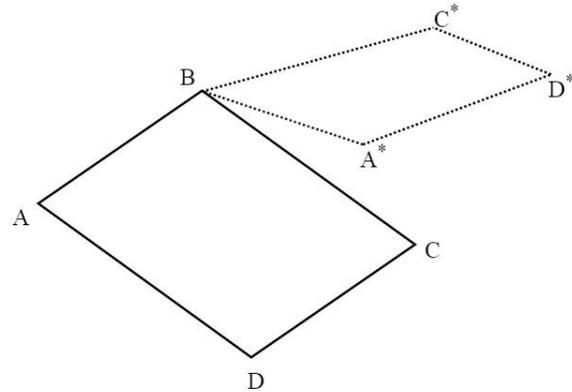


Fig. 2. Figure transformations

Step 8. For each figure F_i find the figure B_i , obtained by rotating F_i around the center of the figure.

Step 9. For each figure F_i find a modified figure R_i .

Step 10. Form a general population P_t , which will contain all the new points created in the previous steps. Thus, $P_t = P_t + T_i + U_i + Q_i + B_i + R_i, i = \overline{1, w}$.

Step 11. For all potential solutions from P_t find the value of the function f and sort the potential solutions from the best to the worst.

Step 12. Leave in P_t only m best solutions and check the fulfillment of the stop criterion.

Step 13. If the stop condition is not met, go to step 3. Otherwise, complete the algorithm and the best element in the population will be considered the best solution. Consider that the stop criterion may be:

- a given number of iterations;
- the average value of fitness function in neighboring populations is less than specified, etc;
- the worst value of fitness function in neighboring populations is less than specified.

THE EXPERIMENTAL RESULTS

Figure 3 shows the experimental results for the Schwefel 2.20 function in 10-dimensional space. The comparative graph of results is presented in Figure 4.

It shows how well-known methods and method of deformed stars (MODS) found a solution during the execution with the stop condition by iterations. As we can see, MODS found a solution on the initial iterations, in contrast to other known methods.



Test function	Global minimum	1000 iterations, 10 launches, accuracy = 10^{-5}					
		GA	ES	DE	MODS-3	MODS-4	MODS-5
Schwefel 2.20	0	0.143	194.1793	79.7148	0.0	0.0	0.0
Worst fitness-function between populations, $\epsilon = 1 \cdot 10^{-10}$							
Schwefel 2.20	0	8.4094	191.1157	283.9126	0.0	0.0	0.0
Average fitness-function between populations, $\epsilon = 1 \cdot 10^{-10}$							
Schwefel 2.20	0	3.6056	244.8562	170.2587	0.0	0.0	0.0

Fig. 3. Comparison of the results of methods under three conditions of completion

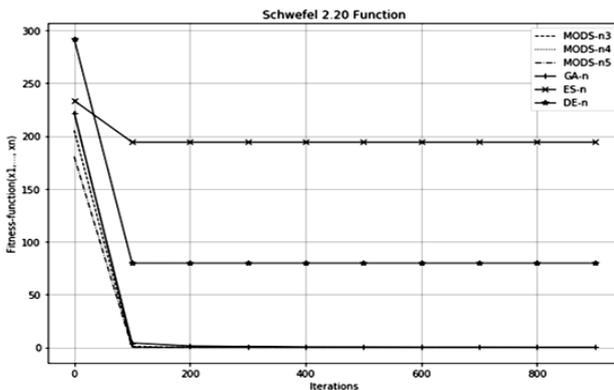


Fig. 4. Convergence plots for the Schwefel 2.20 function, the condition by iterations, coefficient = 2

CHOOOSING THE BEST PARAMETERS FOR MODS

As we can see, first experiments were made with the parameters, which are used to perform transformations in MODS populations, equal to 2, for all stop conditions. Moreover, MODS with these parameters was used for comparison with well-known methods in Figure 3.

To investigate the best configuration of the MODS, it was decided to conduct testing, considering the different values of the input parameters of the method. For the study, it was decided to set the values of all parameters, which are used to perform transformations in MODS populations, equal to 1.5.

After conducting experiments, it was found that even with the change of parameters, MODS was able to find the correct extremes of the tested functions. Thus, we can say that when changing the parameters of the method, the solution will still be found. However, it is important to note, that for co-

efficients = 1.5 the method worked faster. Therefore, the change in the value of the parameter affected the speed of finding the result. The comparative graph of results is presented in Figure 5. It shows how well-known methods and MODS (but, in this case, with parameters, that are equal to 1.5) found a solution with the execution of the stop condition by iterations.

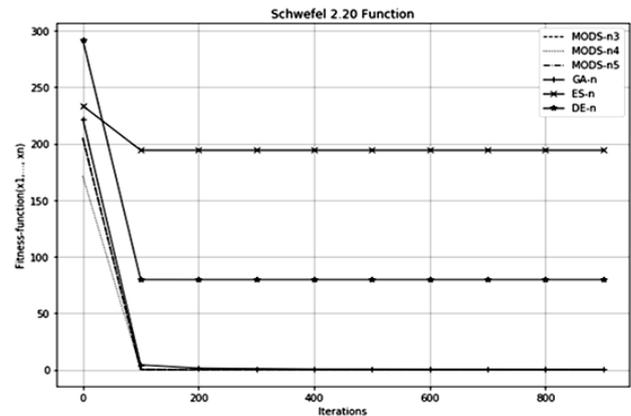


Fig. 5. Convergence plots for the Schwefel 2.20 function, the condition by iterations, coefficients = 1.5

CONCLUSION

The large number of modern practical problems belong to the class of constraint satisfaction problems (CSPs). Such algorithms, as stochastic search, combinatorial optimization methods, and evolutionary algorithms are used to solve these tasks.

We propose to use the idea that underlies all evolutionary algorithms, namely the population approach, which consists of solving problems (1), (2) by evolution the population of potential solutions. The algorithm is called method of deformed stars.

The key elements of the solution search for the problems (1), (2) are special transformations like compressions, and rotations of figures formed by the connection of points, which are the potential solutions. In addition, these groups of points are called deformed stars. MODS belongs to the class of evolutionary algorithms and makes it possible to take into account the surface of the function under study. Its advantages are the speed of convergence and the accuracy of the result compared to other evolutionary methods.

The obtained experimental results allow us to conclude that the proposed method is applicable to solve problems of finding optimal (suboptimal) values, including undifferentiated functions. In implementing the method of deformed stars, significantly fewer steps performed in the wrong direction, in contrast to genetic algorithm, the method of differential evolution and evolutionary strategy as representatives of classic evolutionary paradigm. The



accuracy of the obtained solutions is, on average, higher than that of competing algorithms due to a deeper study of the solution search area.

It is also important to note that in the course of research it was found that when changing the parameters of the MODS solution will still be found.

tion of Systems, Advances in Intelligent Systems and Computing, Springer, Cham, 2019, vol 1019.

[18] M. Antonevych, A. Didyk, V. Snytyuk, "Choice of better parameters for method of deformed stars in n-dimensional case", in *Proc. IT&I*, 2020, pp. 17–20.

Стаття надійшла до редколегії

10.09.2021

REFERENCES

- [1] Vikhar, P. A., "Evolutionary algorithms: A critical review and its future prospects", in *Proc. 2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC)*, 2016, pp. 261–265.
- [2] Arthur E. Carter, Cliff T. Ragsdale, "A new approach to solving the multiple traveling salesperson problem using genetic algorithms", *European Journal of Operational Research*, Volume 175, Issue 1, pp. 246–257, 2006.
- [3] Kierkosz, I., Luczak, M. "A hybrid evolutionary algorithm for the two-dimensional packing problem". *Cent Eur J Oper Res* 22, 2014, pp. 729–753.
- [4] Olariu Stephan, Zomaya Albert Y, *Handbook of Bio-inspired Algorithms and Applications*, Chapman Hall/Crc Computer Information Science, 2006.
- [5] Madsen, S. T. and Widmer, G., "Evolutionary Search for Musical Parallelism", Applications of Evolutionary Computing, in *Proc. EvoWorkshops*, 2005, LNCS 3449 p., pp. 488–497.
- [6] Lawrence J. Fogel, "Evolutionary programming", Alexandria, Virginia, U.S.: *National Science Foundation (NSF)*, 1960.
- [7] Zaineb Chelly Dagdia, Miroslav Mirchev, Chapter 15 – When Evolutionary Computing Meets Astro- and Geoinformatics, Knowledge Discovery in Big Data from Astronomy and Earth Observation, Elsevier, 2020, pp. 283–306.
- [8] R.G.S. ASTHANA, CHAPTER 6 – Evolutionary Algorithms and Neural Networks, In Academic Press Series in Engineering, Soft Computing and Intelligent Systems, Academic Press, 2000, pp. 111–136.
- [9] John Henry Holland, *Adaptation in Natural and Artificial Systems*, Ann Arbor: University of Michigan Press, 1975.
- [10] Rechenberg, Ingo, "Evolutionsstrategie – Optimierung technischer Systeme nach Prinzipien der biologischen Evolution", PhD thesis, Frommann-Holzboog, 1973.
- [11] Hans-Paul Schwefel, "Numerische Optimierung von Computer-Modellen", PhD thesis, 1974.
- [12] Beni, Gerardo, and Jing Wang. "Swarm intelligence in cellular robotic systems." *Robots and biological systems: towards a new bionics?*, Springer, Berlin, Heidelberg, 1993. pp. 703–712.
- [13] Blum, Christian & Li, Xiaodong, *Swarm Intelligence in Optimization*, 2008.
- [14] D.H. Wolpert, W.G. Macready, "No Free Lunch Theorems for Optimization", in *Proc. IEEE Transactions on Evolutionary Computation* 1, 1997, p.67.
- [15] Tmienova N., Snytyuk V., "Method of Deformed Stars for Global Optimization". in *Proc. 2020 IEEE 2nd International Conference on System Analysis & Intelligent Computing (SA-IC)*, 2020, pp. 1–4.
- [16] M. Antonevych, A. Didyk, V. Snytyuk, "Optimization of Functions of Two Variables by Deformed Stars Method", in *Proc. 2019 IEEE International Conference on Advanced Trends in Information Theory (ATIT)*, 2019, pp. 475–480.
- [17] V. Snytyuk, "Method of Deformed Stars for Multi-extremal Optimization. One- and Two-Dimensional Cases", in *Proc. International Conf. Mathematical Modeling and Simula-*



Вибір найкращих параметрів для методу деформованих зірок у n -вимірному просторі

Досліджено задачу оптимізації функції в n -вимірному просторі, яка, у загальному випадку, є поліекстремальною і недиференційованою. Запропоновано новий метод деформованих зірок у n -вимірному просторі. Указаний метод побудований на ідеях і принципах еволюційної парадигми і заснований на припущенні використання потенційних груп розв'язків. Це дозволяє підвищити швидкість та збіжності досягнутого результату. Для оптимізації функції від багатьох змінних використовуються популяції потенційних рішень. Отримано метод, який застосовують для розв'язання задач в n -вимірному просторі, де сукупність розв'язків складається з 3-, 4- та 5-точкових груп. Показано переваги розробленого методу перед генетичним алгоритмом, диференціальною еволюцією та еволюційною стратегією, як найбільш типовими еволюційними алгоритмами. Також проведено експерименти для дослідження найкращої конфігурації параметрів методу деформованих зірок.

Ключові слова: технологія, оптимізація, метод, експеримент.



Марина Антоневич. Студентка факультету інформаційних технологій Київського національного університету імені Тараса Шевченка. Київ, Україна. Сфера наукових інтересів: наука про дані, еволюційні алгоритми, машинне навчання, глибоке навчання, комп'ютерний зір, інтелектуальний аналіз даних.

Maryna Antonevych. Student Faculty of Information Technology, Taras Shevchenko National University of Kyiv. Kyiv, Ukraine. Scientific interests: data science, evolutionary algorithms, machine learning, deep learning, computer vision, data mining.



Анна Дідик. Студентка факультету інформаційних технологій Київського національного університету імені Тараса Шевченка. Київ, Україна. Сфера наукових інтересів: наука про дані, еволюційні алгоритми, веброзробка, вебдизайн, машинне навчання, глибоке навчання.

Anna Didyk. Student Faculty of Information Technology, Taras Shevchenko National University of Kyiv. Kyiv, Ukraine. Scientific interests: data science, evolutionary algorithms, web-development, web-design, machine learning, deep learning.



Наталія Тмєнова. Кандидат фізико-математичних наук, доцент, заступник декана з навчально-виховної роботи. Вища освіта: Київський національний університет імені Тараса Шевченка, магістр соціальної інформатики, викладач математики та інформатики. Київ, Україна. Кандидатська дисертація: 01.05.04 — Системний аналіз і теорія оптимальних рішень, «Граничні теореми для керованих систем». Сфера наукових інтересів: обробка природної мови, технології електронного навчання, дослідження систем масового обслуговування, дослідження випадкових процесів.

Nataliia Tmienova. Candidate of Physical and Mathematical Sciences (PhD), Associate Professor, Deputy Dean of Academic Studies.

Higher Education: Taras Shevchenko National University of Kyiv, Master in Social Informatics, Teacher of Mathematics and Informatics. Kyiv, Ukraine.

Defended Thesis: PhD: 01.05.04 — System analysis and theory of optimal decisions, "The limit theorems for controlled systems".

Scientific interests: natural language processing, e-learning technologies, queuing systems research, random processes research.



Віталій Снитюк. Доктор технічних наук, професор, декан факультету інформаційних технологій.

Вища освіта: Київський національний університет імені Тараса Шевченка. Київ, Україна. Математик, викладач. Кандидатська дисертація: 05.13.12 — Системи автоматизації проектних робіт, «Методи зменшення невизначеності на початкових етапах проектування систем із змінною структурою». Докторська дисертація: 05.13.06 — Інформаційні технології, «Еволюційні технології прийняття рішень в умовах невизначеності». Сфера наукових інтересів: прийняття рішень в умовах невизначеності, обчислювальний інтелект, еволюційне моделювання.

Vitaliy Snytyuk. Doctor of Technical Science, Professor, Dean of the Faculty of Information Technology.

Higher education: Taras Shevchenko National University of Kyiv. Kyiv, Ukraine. Mathematician, Lecturer.

Defended Theses: PhD 05.13.12 — Systems of automation of design works, "Methods to reduce uncertainty on the initial stages of systems design with variable structure". Doctoral 05.13.06 — Information technology, "Evolutionary technologies of decision making under uncertainty".

Scientific interests: decision making under uncertainty, computational intelligence, evolutionary modelling.



МАТЕМАТИЧНІ ОСНОВИ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ



УДК 004.942

DOI: <https://doi.org/10.17721/AIT.2021.1.04>

М. В. Пирог, orcid.org/0000-0003-2588-6066,
Київський національний університет імені Тараса Шевченка, Київ, Україна,
І. І. Гарко, orcid.org/0000-0003-0671-6336,
Національний технічний університет України
“Київський політехнічний інститут імені Ігоря Сікорського”, Київ, Україна,
К. К. Духновська, orcid.org/0000-0002-4539-159X,
Київський національний університет імені Тараса Шевченка, Київ, Україна

Моделювання термофізіологічного стану людини

Наведено аналіз математичних моделей, які можна використати для прогнозування термофізіологічного стану людини в різних умовах навколишнього середовища у процесі розроблення інформаційних систем її життєзабезпечення. Базисним елементом всіх розглянутих моделей є рівняння теплового балансу. Відповідно до цього рівняння сумарна тепловіддача організму має дорівнювати його теплопродукції. Усі величини рівняння теплового балансу можуть бути визначені спеціальними експериментами або за допомогою аналітичних методів, розроблених на основі класичної теорії термодинаміки. Розглянуто континуальні та дискретно-судинні моделі, проаналізовано їхні переваги та недоліки. Континуальні моделі є, по суті, спрощеним записом біотеплових рівнянь. У цих моделях впливом кровотоку кожної окремої судини нехтують, кровопостачання осереднюється за досліджуванним об'ємом. Дискретно-судинні моделі є сукупністю біотеплових рівнянь, що описують кровоток у кожній окремій судині. Дискретно-судинні моделі термофізіологічного стану людини нині малозастосовні, що пов'язано зі складною і не досить дослідженою геометрією судин. Більше використовують континуальні моделі, для побудови яких застосовують мультикомпаратментальний підхід.

У результаті у статті наведено порівняльну таблицю континуальних моделей та інформаційних систем, які використовують ці моделі.

Ключові слова: термофізіологічний стан людини, теплообмін, математичне моделювання, інформаційні системи.

Для цитування (for citation): М. В. Пирог, І. І. Гарко, К. К. Духновська. “Моделювання термофізіологічного стану людини,” *Сучасні інформаційні технології*, vol. 1, p. 30–38, 2021.

ВСТУП

Усі функціональні системи організму людини тією чи іншою мірою мають захисну реакцію на впливи зовнішнього середовища. Розвиток медичної кібернетики, прикладної математики та обчислювальних методів створили передумови для побудови інформаційних технологій для аналізу стану організму під впливом зовнішнього середовища. Математичні моделі функціональних систем організму та їхній комп'ютерний аналіз, безумовно, доповнюють систему експериментальних і клінічних даних, що отримується традиційними для фізіології і медицини методами.

За останні півстоліття зусиллями співдружності математиків, фізиків і фізіологів побудовано марематичні моделі практично всіх

досліджуваних функціональних систем організму людини. Дослідженню різних аспектів моделювання стану людини присвячено роботи таких авторів: І. Й. Єрмакова, К. Г. Лябах, F. d'Ambrosio, V. Candas, Fiala, D. L. Fogt, C. Huizenga, S. Tanabe.

Однією з актуальних проблем у галузі розроблення систем життєзабезпечення є забезпечення теплового комфорту для людини в межах системи «людина – довкілля», зокрема, це відноситься до задачі захисту людини від переохолодження, перегріву або зневоднення. Експериментальні дослідження в такому широкому діапазоні параметрів навколишнього середовища по фазовому стану, температурі й тиску вкрай трудомісткі і часто пов'язані з небезпекою для здоров'я та життя його учасників.

© Пирог М. В., Гарко І. І., Духновська К. К., 2021



Одним із варіантів розв'язання цієї задачі є розроблення моделей процесів теплообміну в системі «людина – довкілля», які дозволяють проводити широкий комплекс досліджень без експериментів над людьми.

Кожна модель функціональних систем описується властивими їй специфічними параметрами, які оцінюють стан указаної функціональної системи. У цій статті проведено аналіз термофізіологічних моделей, оскільки температура тіла людини є основним показником її здоров'я.

РІВНЯННЯ ТЕПЛООВОГО БАЛАНСУ ЛЮДИНИ ЯК БАЗИСНИЙ ЕЛЕМЕНТ МОДЕЛЕЙ

Регульовану постійність внутрішнього середовища живого організму називають гомеостазом [1]. Термін характеризує динамічну сталість у підтримці різноманітних констант організму: температури тіла, складу крові, кров'яного тиску тощо. До найдосконаліших гомеостатичних механізмів в організмі людини належать процеси терморегуляції. Сталість температури може зберігатися лише за умови рівності теплоутворення і тепловіддачі. Це досягається за допомогою фізіологічних механізмів теплообміну. Організм людини є складною біологічною саморегулюючою системою, що перебуває у стані безперервного теплового обміну із середовищем, в якому він міститься, і до цієї системи можуть бути застосовані перший і другий закони термодинаміки. Для збереження сталої температури тіла вся система має перебувати в термостабільному стані. Для визначення рівності між теплоутворенням в організмі та тепловіддачею використовують термін «тепловий баланс».

Теплового балансу досягають організацією і управлінням процесів, спрямованих на утворення теплоти в організмі (телопродукція) і його виведення (тепловіддача) [2].

Нехай Q – це кількість внутрішньої енергії (теплоти), яку має тіло за даної температури T . Цю енергію прийнято називати ентальпією.

Відповідно до першого закону термодинаміки (закону збереження енергії) сумарна тепловіддача організму повинна дорівнювати його теплопродукції, тому рівняння теплового балансу є таким [2]:

$$Q = M - Q^R + Q^C + Q^K + Q^b + Q^E, \quad (1)$$

де Q – зміна температурного поля організму відносно комфортного рівня; M – метаболізм людини; Q^R – втрата теплоти радіацією; Q^C – втрата теплоти конвекцією; Q^K – втрата теплоти кон-

дукцією; Q^b – перенесення теплоти кров'ю; Q^E – втрата теплоти випаровуванням.

Усі величини рівняння теплового балансу можуть бути визначені спеціальними експериментами або за допомогою аналітичних методів, розроблених на основі класичної теорії термодинаміки.

До одного із чинників вироблення теплоти в організмі людини належить метаболізм. Він відповідає за підтримання сталої температури тіла людини, здійснення роботи, росту і розвитку організму та забезпечення структури і функцій усіх клітинних елементів.

У результаті окиснювально-відновлювальних реакцій в організмі людини відбувається перетворення речовин, багатих енергією, на речовини, що містять менший енергетичний потенціал, зі звільненням енергії в тій чи іншій формі [3]. Метаболізм у різних тканинах проходить із різною швидкістю. Швидкість метаболізму також залежить від віку людини.

У навколишнє середовище людина віддає теплоту різними шляхами: радіацією, конвекцією, випаровуванням, диханням. Тепловіддача радіацією залежить від температури тіла, одягу та навколишніх поверхонь, їхньої площі тощо. Причому теплообмін може йти як із позитивним, так і з негативним тепловим балансом для людини. Позитивний радіаційний тепловий баланс спостерігається, коли середня температура поверхонь, що оточують людину, вища температури поверхні її тіла. У цьому випадку за рахунок інфрачервоного випромінювання тіло людини нагрівається. Негативний радіаційний тепловий баланс спостерігається, коли середня температура поверхонь, що оточують людину, нижча температури поверхні його тіла. Тепловіддача радіацією може бути визначена за різними рівняннями: Стефана – Больцмана, Вітте, Фагнера. Для практичних розрахунків використовують рівняння Стефана – Больцмана [3]:

$$Q^R = h^R A_{\text{рад}} (T_1 - T_2), \quad (2)$$

де h^R – коефіцієнт тепловіддачі радіацією; $A_{\text{рад}}$ – площа поверхні тіла, що бере участь у радіаційному теплообміні; T_1, T_2 – температури тіл, що беруть участь у радіаційному теплообміні.

Тепловіддача конвекцією визначається на основі закону Ньютона – Ріхмана за рівнянням [3]:

$$Q^C = h^C A_C (T_1 - T_2), \quad (3)$$

де h^C – коефіцієнт тепловіддачі конвекцією; A_C – площа поверхні тіла, що бере участь у конвекційному теплообміні; T_1, T_2 – температури



тіл, що беруть участь у конвекційному теплообміні.

Тепловіддача кондукцією визначається за законом Фур'є [4]:

$$Q^K = \lambda \frac{T_1 - T_2}{\sigma} S \tau, \quad (4)$$

де λ – коефіцієнт теплопровідності; T_1, T_2 – температура внутрішньої і зовнішньої сторін пакетів матеріалів; σ – товщина тканини; S – площа поверхні, що стикається з іншою поверхнею; τ – час стикання.

У звичайних умовах втрати теплоти кондукцією невеликі, оскільки невелика площа дотику тіла з іншими поверхнями. Людина втрачає теплоту кондукцією з поверхні підошов, площа яких становить 3 % усієї поверхні тіла. Унаслідок цього тепловіддача кондукцією не має великого значення і нею можна знехтувати.

Втрати теплоти випаровуванням мають велике значення в умовах перегрівання організму і в разі виконання людиною фізичної роботи [3]:

$$Q^E = Q_{\text{вип.диф.}} + Q_{\text{вип.дих.}} + Q_{\text{вип.п.}} + Q_{\text{дих.н.}}, \quad (5)$$

де $Q_{\text{вип.диф.}}$ – втрата теплоти внаслідок випаровування дифузійної вологи з поверхні шкіри; $Q_{\text{вип.дих.}}$ – втрата теплоти внаслідок випаровування вологи з верхніх дихальних шляхів; $Q_{\text{вип.п.}}$ – втрата теплоти внаслідок випаровування поту; $Q_{\text{дих.н.}}$ – втрата теплоти під час нагрівання повітря, що вдихається.

Тепловіддача, особливо за високої температури повітря, може здійснюватися шляхом випаровування дифузійної вологи з поверхні шкіри та верхніх дихальних шляхів. Швидкість випаровування вологи з поверхні тіла залежить від різниці парціальних тисків пари у прикордонному шарі близько шкіри та в навколишньому повітрі; швидкості повітря; повітро- і паропроникнення одягу; площі поверхні, зволоженої потом. Втрати теплоти шляхом випаровування дифузійної вологи з поверхні шкіри можна визначити за рівнянням [3]:

$$Q_{\text{вип.диф.}} = 3.06 \cdot 10^{-3} S (256T_{\text{ш}} - 3360 - p_a), \quad (6)$$

де S – площа поверхні тіла людини; $T_{\text{ш}}$ – температура шкіри; p_a – парціальний тиск, який приймають рівним 2660 Па.

Повітря, що вдихається, у процесі проходження по дихальних шляхах зволожується за рахунок випаровування вологи зі слизового шару. Під час виходу водяна пара частково конденсується, тому повітря, що видихається, містить більше вологи, ніж те, що вдихається. Втрати теплоти при випаровуванні вологи з верхніх дихальних шляхів можна визначити з рівняння [3]:

$$Q_{\text{вип.д.}} = 14.9 \cdot 10^{-6} Q_{\text{тп}} (5880 - p_a), \quad (7)$$

де $Q_{\text{тп}}$ – теплопродукція людини.

Величина потовиділення багато в чому визначається рівнем фізичної активності, умовами та ступенем відповідності одягу умовам зовнішнього середовища. Основними показниками навколишнього середовища, що впливають на величину тепловтрат за рахунок випаровування поту, є відносна й абсолютна вологість. Відносна вологість визначає інтенсивність випаровування вологи з поверхонь, зокрема і з поверхні людського тіла, та визначається таким чином [3]:

$$\varphi = \frac{e}{E_{\text{ф}}} \cdot 100\%, \quad (8)$$

де e – напруга водяної пари в повітрі (абсолютна вологість); $E_{\text{ф}}$ – максимально можлива напруга водяної пари при температурі повітря.

Величину потовиділення (втрати теплоти за рахунок випаровування поту) можна визначити за формулою Н. К. Вітте [7]:

$$Q_{\text{вип.}} = 17.3 (E_{\text{ф}} - e) (0.5 + \sqrt{V_{\text{п}}}), \quad (9)$$

де $V_{\text{п}}$ – об'єм поту.

Комфортний рівень тепловіддачі випаровуванням $Q_{\text{вип.к.}}$, визначається з рівняння [3]:

$$Q_{\text{вип.к.}} = 0.36S \left(\frac{Q_{\text{тп}}}{S} - 58 \right), \quad (10)$$

де S – площа поверхні тіла людини.

Втрати теплоти диханням визначають за рівнянням Фігнера [3]:

$$Q_{\text{дих.}} = 0.0012 Q_{\text{ет}} (34 - T_{\text{п}}), \quad (11)$$

де $Q_{\text{ет}}$ – енерговитрати людини; $T_{\text{п}}$ – температура повітря.

Розраховуючи втрати теплоти диханням слід приймати температуру повітря, що видихається: при високій температурі навколишнього середовища – 36 °С; при середній температурі навколишнього середовища – 34 °С; при низькій температурі навколишнього середовища – 30 °С.

Отже, головними чинниками температури тіла людини є метаболізм, конвекція, кондукція, радіація і втрати теплоти диханням і випаровуванням.

МУЛЬТИКОМПАРТМЕНТАЛЬНІ МОДЕЛІ ТЕПЛООБМІНУ ЛЮДИНИ

Майже всі інформаційні технології моделювання термофізіологічного стану людини використовують мультикомпаратментальну модель теплообміну людини.

За означенням, наведеним у [4], компартмент – це певна кількість речовини, що виділяється в біологічній системі та має властивість єдності, тому у процесах транспортування і хімічних перетворень її можна розглядати як одне ціле. Моделі, в яких досліджувана система представляється у вигляді сукупності компартментів, називають мультикомпаратментальними.

Клас мультикомпаратментальних динамічних моделей теплообміну і терморегуляції людини



описує теплопродукцію в органах і тканинах, теплоперенесення кровотоком, кондукцію, конвекцію, випромінювання, випаровування зі шкіри та верхніх дихальних шляхів, аферентні та еферентні терморегуляторні процеси.

В основі структури моделей міститься компартмент. Він має джерело енергії, здатність приймати й віддавати енергію іншим компартментам та обмінюватися нею із середовищем. Набір компартментів і формування відношень між ними задають ступінь апроксимації тіла людини. Кожен компартмент відповідає певній частині тіла або тканині організму. Апроксимація тіла людини може проводитися на довільну кількість геометричних частин.

АНАЛІЗ МОДЕЛЕЙ ТЕРМОФІЗІОЛОГІЧНОГО СТАНУ ЛЮДИНИ

Завдання моделювання термофізіологічного стану людини зводиться до розроблення математичної моделі теплового режиму з урахуванням усіх присутніх факторів середовища, специфіки професійної діяльності та фізіології людини.

Теплообмін у живій тканині виключає точне математичне моделювання, оскільки є складним процесом, що містить теплопровідність у тканинах, конвекцію та перфузію крові, а також генерацію метаболічної теплоти. Нині для опису теплопередачі всередині живих тканин розроблено багато математичних моделей, що використовуються в різних галузях біології та медицини. Однак основою побудови більшості відомих моделей перенесення біотеплоти є два базові підходи: континуальний і дискретно-судинний. Континуальні моделі є, по суті, спрощеним записом біотеплових рівнянь. У цих моделях впливом кровотоку кожної окремої судини нехтують, кровопостачання усереднюється за досліджуваним об'ємом. Дискретно-судинні моделі є сукупністю біотеплових рівнянь, що описують кровоток у кожній окремій судині та дозволяють прогнозувати розподіл температури [4].

Найбільшого поширення набула дискретно-судинна модель васкуляризованої тканини, запропонована в [5]. Модель базується на чотирьох припущеннях:

- теплообміном між кров'ю та тканиною в преартеріолах і поственулах нехтують;
- потік крові у дрібних капілярах вважають ізотропним, напрямком кровотоку нехтують;
- великі кровоносні судини в безпосередній близькості від капілярного русла не вносять вклад в обмін енергією між тканиною та капілярною кров'ю (іншими словами, ця модель не враховує локальну геометрію судин);
- температура крові в артеріолах, що постають капілярне русло, дорівнює температурі ядра тіла [5].

Обмін енергією відбувається миттєво: температура крові врівноважується температурою тканини, базуючись на цих припущеннях, у [6] змодельовано вплив крові як ізотропного джерела теплоти, пропорційного швидкості потоку крові та різниці між температурою тіла і локальною температурою тканини. По суті, модель, яку описано в [5], є класичним біотепловим рівнянням, що враховує вплив кровотоку та метаболічної теплоти на баланс енергії у тканині (рівняння Пеннеса):

$$\rho_{ti} C_{ti} \frac{dT_{ti}}{dt} =$$

$$= \nabla(k_{ti} \nabla T_{ti}) + \rho_b C_b W_b (T_{art} - T_{ti}) + q_m, \quad (12)$$

де ρ_{ti} – густина тканини; C_{ti} – теплоємність тканини; T_{ti} – температура тканини; k_{ti} – теплопровідність тканини; ρ_b – щільність крові; C_b – теплоємність крові; W_b – швидкість перфузії крові; T_{art} – температура артеріальної крові; q_m – метаболічна теплота.

Генерацію метаболічної теплоти, що розглядається в цій моделі, передбачають однорідно розподіленою по всій тканині, перфузію крові також вважають однорідною та ізотропною. Згідно з моделлю [5] теплова рівновага виникає безпосередньо в капілярній ланці мікроциркуляторного русла (кров при артеріальній температурі надходить у капіляри, де відбувається теплообмін, температура знижується до температури тканини). Температура, з якою кров входить у венозний кровообіг, дорівнює локальній температурі тканини [5].

Однак, варто звернути увагу на недоліки такої моделі:

- теплообмін між кров'ю та тканиною відбувається не лише в капілярах, навпаки, він спостерігається також у преартеріолярних та поственулярних судинах, що мають діаметри в діапазоні 70–500 мкм;
- напрямком кровотоку є важливим фактором, що впливає на обмін енергією між судинами та тканиною, однак рівняння Пеннеса не враховує цей чинник;
- рівняння Пеннеса не враховує також локальну геометрію судин (обмін енергією з великими судинами, передачу теплоти у протиток між парами артерія-вена та розгалуження-звуження судин);
- температура артеріальної крові безперервно змінюється від аорти до артерій та в артеріолах, а також далі в системі венозного відтоку.

Таким чином, температура крові в артеріолах не дорівнює температурі ядра тіла, а температура венозної крові не дорівнює локальній температурі тканини. Зазначені наближення переоцінюють



вплив кровотоку на локальну температуру тканини. Нехтування впливом різної направленості потоків крові в [5] призводить до суттєвих помилок під час обчислення результатів.

Для виправлення зазначеного недоліку в [7] припущено, що поле теплообміну всередині тканини має бути змодельоване на основі природних умов судинної анатомії. З урахуванням просторових і часових варіацій швидкості (v), джерела метаболічної теплоти (q_m), сталості

$$\rho_{ti,eff} C_{ti,eff} \frac{\partial T_{ti}}{\partial t} = \nabla(k_{ti,eff} \nabla T_{ti}) - (\rho_b C_b W^* (T_{art} - T_{ti}) - \rho_b C_b v \nabla(T_{ti}) + \nabla(k_p \nabla T_{ti})) + q_m, \quad (14)$$

де k_p – провідність перфузії,

$$\rho_{ti,eff} = (1 - \varepsilon_b) \rho_{ti} + \varepsilon_b \rho_b, \quad C_{ti,eff} = (1 - \varepsilon_b) C_{ti} + \varepsilon_b C_b, \quad k_{ti,eff} = (1 - \varepsilon_b) k_{ti} + \varepsilon_b k_b,$$

тут ε_b – пористість тканини, через яку рухається кров.

Модель, що описана в [8], досить складно застосувати на практиці, оскільки потрібна наявність детальних відомостей про судинну мережу та перфузію крові. У моделі, так само як у всіх континуальних моделях, найбільша увага приділяється теплообміну між кровоносною судиною та навколишньою тканиною. З огляду на різниці температур і близьке розташування венозних і артеріальних кровоносних судин, масмо також і теплообмін у зустрічному потоці, але це у моделі не враховується.

У [9] представлено нову судинну біотеплову модель, в якій береться до уваги зустрічний потік крові. Цю модель отримано на основі припущення, що невеликі артерії та вени паралельні, а напрямок потоків протилежний. Таке припущення може бути застосовано, в основному, для проміжних шарів шкіри. В анатомічному дослідженні, проведеному на кінцівках кролика, в [9] визначено три судинних шари (глибокий, середній і шкірний) у зовнішньому шарі тканини глибиною 1 см.

Припущення для моделі, описаної в [12], такі:

- лімфатичною втратою рідини нехтують, таким чином, масові витрати в артерії та у вени рівні;
- кровоток просторово однорідний;
- теплопередача у площині, перпендикулярній до пари артерія-вена, більша, ніж уздовж судин;
- залежність температури в радіальному напрямку пари артерія-вена є лінійною;
- температура межі артерія-вена дорівнює середньому значенню температур артерії та вени;
- температура крові на виході з відвідних капілярів та на вході у вени дорівнює температурі венозної крові.

Модель, що подається у [12], така:

$$\rho_{ti} C_{ti} \frac{dT_{ti}}{dt} = \nabla(k_{eff} \nabla T_{ti}) + q_m, \quad (15)$$

де k_{eff} – ефективна теплопровідність, яка обчислюється так:

фізичних властивостей тканини біотеплове рівняння у [7] записано так:

$$\rho_{ti} C_{ti} \frac{dT_{ti}}{dt} = k_{ti} \nabla^2 T_{ti} + \rho_b C_b v \nabla(T_{ti}) + q_m. \quad (13)$$

Серед дискретно-судинних біотеплових моделей модель [8–10] є найбільше відпрацьованою. У [8] зроблено припущення, що загальний контрольний об'єм тканини доцільно розділити на підоб'єм обезкровленої тканини та підоб'єм крові, у результаті отримано рівняння:

$$k_{eff} = k_{ti}(1 + Pe_i V(\zeta)),$$

де $\zeta = \frac{x}{L}$ – безрозмірна відстань, що визначається; L – товщина шару тканини; $V(\zeta)$ – безрозмірна функція судинної геометрії, яка може бути обчислена, якщо є відомості про судини; Pe_i – число Пекле, яке визначається за такою формулою:

$$Pe_i = \frac{2\rho_b C_b r_i v_i}{k_b},$$

де r_i – радіус судини;

v_i – швидкість крові на вході в шар тканини.

Така модель вимагає багато докладних відомостей про анатомію та судинну геометрію, наприклад, щільність судини, розмір і відстань між артерією та веною для кожної судини цієї пари, а також швидкість перфузії крові. Це все ускладнює застосування такої моделі.

Серед континуальних моделей нині досить відомою є модель Fiala [13]. Відповідно до цієї моделі, серед складових системи виокремлюють такі: мозок, шия, грудна клітка, живіт, кінцівки. Тут кожен елемент складається з шарів: шкіра, жир, м'язи та кістки. Рівняння теплового балансу, що застосовується у цій моделі, має вигляд

$$k \left(\frac{\partial^2 T}{\partial^2 r} + \frac{w}{r} \frac{\partial T}{\partial r} \right) + q_m + \rho_b w_b C_b (T_{b,art} - T) = \rho C \frac{\partial T}{\partial t}. \quad (16)$$

У моделі беруть участь потік крові, метаболічні реакції та місцевий теплообмін між людиною і середовищем [13].

У [14] наведено універсальну модель із зосередженими параметрами. Вона теж належить до дискретно-судинних моделей. У ній враховано потік крові, дихання, випаровування, метаболічні реакції, фізичне навантаження, пітливість, тремтіння, серцевий викид і місцевий теплообмін між людиною та середовищем. У



моделі також передбачено врахування наявності одягу та захисного спорядження. Тіло людини у цій моделі апроксимується заданою кількістю компартментів залежно від мети використання моделі для розв'язання задач із дослідження фізіології людини. На основі моделі із зосередженими параметрами створено модель для оцінювання впливу захисного спорядження на термофізіологічний стан людини [15], модель для оцінювання впливу електромагнітного поля радіочастотного й ультразвукового діапазону на гіпертермію [16], модель для визначення рівня температурного комфорту людини у приміщенні [17], модель для прогнозування теплового стану людини під час перебування у воді [18].

У роботах [14–18] для побудови математичної моделі кожен компартмент апроксимується матеріальною точкою з характеристиками цього компартмента, тобто характеристики компартмента «зосереджуються» в одній точці. У результаті отримуємо задачу Коші:

$$c_{ij}m_{ij} \frac{dT_{ij}}{dt} + \Delta T_{ij} = M_{ij}^* + M_{ij}^{ph} + M_{ij}^{sh} + Q_{ij-1}^K - Q_{ij}^K - Q_{ij}^b - Q_{ij}^{air} \quad (17)$$

$$V_b \rho_b c_b \frac{dT_b}{dt} = \sum_{i=1}^N \sum_{j=1}^Z w_{ij} \rho_b c_b T_{ij} - W \rho_b c_b T_b - Q^{RS},$$

де c – питома теплоємність; m – маса; T – температура; t – час; M – метаболізм; Q – тепловий потік; V – об'єм; W – серцевий викид.

Індекси:

ij – індекси компартментів; $*$ – початкове значення; b – кров; ph – фізична активність; RS – теплові втрати з диханням; sh – холодове тремтіння скелетних м'язів; K – кондукція; air – повітря.

До плюсів моделі потрібно віднести швидкість обчислень усіх температур компартментів людського тіла. Проте така модель не надає розподілу температур тіла людини у просторі, що не є правильним, вона дає лише середні температури компартментів. Щоб увести нове джерело теплоти, потрібно в цій моделі додавати новий компартмент, що призводить до модифікації інформаційної системи в цілому.

У [19] розглянуто модель із розподіленими параметрами. Температура тіла людини неоднакова на різних його ділянках. Тому в [19] введено температурне поле $T(x, y, z, t)$ – динамічна сукупність значень температури у всіх точках тіла людини. Швидкість зміни температури задано вектором:

$$\nabla(T) = \left(\frac{\partial T}{\partial x}, \frac{\partial T}{\partial y}, \frac{\partial T}{\partial z} \right),$$

де $\nabla(T)$ – градієнт температури.

У результаті отримано систему диференціальних рівнянь у частинних похідних для прогнозування стану людини в екстремальних умовах:

$$c_{ij}m_{ij} \frac{dT_{ij}}{dt} + \Delta T_{ij} = M_{ij}^* + M_{ij}^{ph} + M_{ij}^{sh} + Q_{ij-1}^K - Q_{ij}^K - Q_{ij}^b - Q_{ij}^{air}, \quad (18)$$

$$V_b \rho_b c_b \frac{\partial T}{\partial t} - \lambda_{ij} \Delta T = \sum_{i=0}^{N-1} \sum_{j=0}^{L-1} w_{ij} \rho_b c_b T_{ij} - W \rho_b c_b T_b - Q^{RS},$$

де Δ – оператор Лапласа.

Для розв'язання цієї системи необхідні додаткові умови: початкові та граничні.

Початкові умови визначають температури кожного компартмента та кожного шару в початковий час:

$$\text{при } t = t_0, T_{ij} = T_{ij}^0.$$

Для розв'язання цієї системи використовують граничні умови четвертого роду. Для сусідніх шарів умови повинні виконуватись на межі між ними. Наприклад: межа між циліндром, що апроксимує внутрішні органи, та циліндром, що апроксимує м'язи, одна й та сама та належить як до першого циліндру, так і до другого. Можемо написати, що температура або її градієнт на межі циліндра, що апроксимує внутрішні м'язи, дорівнює температурі, або її градієнту на межі циліндра, що апроксимує м'язи, тобто,

якщо $x \in G_x^{ij}$:

$$T_{xij} = T_{xij-1}.$$

$$\frac{\partial T_{i,j}}{\partial x} = \frac{\partial T_{i,j-1}}{\partial x}, \quad \forall t \geq 0, j \neq sk;$$

якщо $y \in G_y^{ij}$:

$$T_{yij} = T_{yij-1}.$$

$$\frac{\partial T_{i,j}}{\partial y} = \frac{\partial T_{i,j-1}}{\partial y}, \quad \forall t \geq 0, j \neq sk;$$

якщо $z \in G_z^{ij}$:

$$T_{zij} = T_{zij-1}.$$

$$\frac{\partial T_{i,j}}{\partial z} = \frac{\partial T_{i,j-1}}{\partial z}, \quad \forall t \geq 0, j \neq sk;$$

де G_x^{ij} , G_y^{ij} , G_z^{ij} – проєкції границі i -го компартмента j -го шару на координатні площини.

Для температури на границях між шарами, що апроксимують шкіру людини, і шаром повітря має виконуватися закон Фур'є:

для $j = sk, x \in G_x^{i,sk}$:

$$\frac{\partial T_{i,sk}}{\partial x} = a_{ij} h_{i,sk}^K (T_{i,sk} - T^{air}), \quad \forall t \geq 0;$$

для $j = sk, z \in G_y^{i,sk}$:

$$\frac{\partial T_{i,sk}}{\partial y} = a_{ij} h_{i,sk}^K (T_{i,sk} - T^{air}), \quad \forall t \geq 0;$$

для $j = sk, x \in G_z^{i,sk}$:

$$\frac{\partial T_{i,sk}}{\partial z} = a_{ij} h_{i,sk}^K (T_{i,sk} - T^{air}), \quad \forall t \geq 0.$$



У результаті отримано математичну динамічну модель термofізіологічного стану людини, яка може надати розподіл температур тіла людини у просторі.

Наведемо порівняльні характеристики моделей терморегуляції людини (таблиця).

ТАБЛИЦЯ.
ПОРІВНЯЛЬНА ХАРАКТЕРИСТИКА
МОДЕЛЕЙ ТЕРМОРЕГУЛЯЦІЇ ЛЮДИНИ

Модель	Модель Fiala	Із зосередженими параметрами	Із розподіленими параметрами
Кількість компартментів	187	39	39
Дані, з якими може працювати модель	Електронний манекен	Електронний манекен	Дані реальної людини
Важелі впливу на зміни температур, які враховано	Потік крові, метаболічні реакції та місцевий теплообмін між людиною й середовищем, фізичне навантаження, одяг	Потік крові, дихання, випаровування, метаболічні реакції, фізичне навантаження, пітливість, тремтіння, серцевий викид, місцевий теплообмін між людиною й середовищем, одяг або захисне спорядження.	Потік крові, дихання, випаровування, метаболічні реакції, фізичне навантаження, пітливість, тремтіння, серцевий викид, місцевий теплообмін між людиною і середовищем, одяг або захисне спорядження.
Розподіл температур у просторі	Так	Ні	Так
Точність	1,63 °С	0,204 °С	1,3 °С
Час прогнозування	Відомості відсутні	Менше 1 хв	10 хв

ВИСНОВКИ

Проведено порівняння кількох математичних моделей для оцінювання стану терморегуляції в організмі. Математичне моделювання процесів терморегуляції реалізовано у програмній системі й експериментально досліджено на спеціалізованому манекені.

Представлені в поданій праці результати доводять ефективність розробленої математичної моделі, що відкриває можливості для подальшої роботи з її вдосконалення, розширення наборів експериментальних даних для аналізу та підвищення ефективності оцінювання стану організму.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

[1] Я. Гонський, Т. Максимчук, М. Калинський *Біохімія людини*, Тернопіль: Укрмедкнига, 2002.

[2] C. W. Sheppard, "The mathematical basis of the interpretation of tracer experiments in closed steady-state systems", *Journal of Applied Physics*, vol. 22, № 4, pp. 510–520, 1951.

[3] C. H. Wyndham, A. R. Atkins A physiological scheme and mathematical model of temperature regulation in man. *Pflügers Archiv*, vol. 303, pp. 14–30, 1968.

[4] И. Ермакова, «Механизмы физиологической терморегуляции у человека», автореф. дис. на соискание научной степени д-ра биол. наук, Киев, 1989.

[5] H. N. Pennes, "Analysis of tissue and arterial blood temperatures in the resting forearm", *Journal of Applied Physiology*, vol. 1, № 2, pp. 93–122, 1948.

[6] E. H. Wissler, "Pennes 1948 paper revisited", *Journal of Applied Physiology*, vol. 85, № 1, pp. 35–41, 1998.

[7] H. G. Klinger, "Heat transfer in perfused biological tissue", *General theory. Bulletin of Mathematical Biology*, vol. 36, pp. 403–415, 1974.

[8] M. M. Chen, K. R. Holmes, "Microvascular Contributions in Tissue Heat Transfer", *Ann. N. Y. Acad. Sci.*, vol. 335, pp. 137–150, 1980.

[9] S. Weinbaum, L. M. Jiji, D. E. Lemons, "Theory and experiment for the effect of vascular microstructure on surface tissue heat transfer. Part I. Anatomical foundation and model conceptualization", *ASME Journal of Biomechanical Engineering*, vol. 106, pp. 321–330, 1984.

[10] S. Weinbaum, L. M. Jiji, "A new simplified bioheat equation for the effect of blood flow on local average tissue temperature", *ASME Journal of Biomechanical Engineering*, vol. 107, pp. 131–139, 1985.

[11] A. Bhowmik, "Conventional and newly developed bioheat transport models in vascularized tissues", *Journal of Thermal Biology*, vol. 38, № 3, pp. 107–125, 2013.

[12] А. А. Сагайдачный, "Восстановление спектра колебаний кровотока из спектра колебаний температуры пальцев рук, дисперсия температурного сигнала в биоткани", *Регионарное кровообращение и микроциркуляция*, №1, с. 76–82, 2013.

[13] K. Kubaha, D. Fiala, J. Toftum, Human projected area factors for detailed direct and diffuse solar radiation analysis. *International Journal of Biometeorology*, vol.49(2), pp. 113–129, 2004.

[14] І. Й. Ермакова, *Стан та перспективи розвитку інформатики в Україні*, Київ: Наукова думка, 2010.

[15] I.Yermakova, K.Dukhnovskaya, A.Nikolaienko, O.Troynikov, N.Nawaz, "Influence of exercise intensity on thermophysiological responses of firefighters wearing different firefighters protective clothing ensembles", *5th ESPC and Nokobetef 10. Future of protective clothing*, 2012. Pp. 75.

[16] И. И. Ермакова, Ю. П. Тадеева, Н. Г. Иванушкина, "Эффект региональной электромагнитной гипертермии: результаты моделирования", *Электроника и связь. Тематический выпуск "Проблемы электроники"*, № 1, с. 132–136, 2007.

[17] И. И. Ермакова, "Динамическая модель оценки температурного комфорта человека", *Электроника и связь. Тематический выпуск "Проблемы электроники"*, № 2, С. 81–85, 2008.

[18] А. Ю. Ніколаєнко, "Інформаційна технологія прогнозування термofізіологічного стану людини під час фізичної активності в різних середовищах", дисертація на здобуття наукового ступеня кандидата технічних наук, НАН України, Київ, 2019.

[19] Dukhnovska K. The boundary value problem for the heat transfer task between a human and the environment. *Фізико-математичне моделювання та інформаційні технології*. вип. 30, стор. 29–40, 2020.

Стаття надійшла до редколегії

16.08.2021



Modeling of thermophysiological state of person

This article presents an analysis of mathematical models that can be used to predict the thermophysiological state of man in different environmental conditions in the development of information systems for its life support. The basic element of all considered models is the equation of heat balance. According to this equation, the total heat transfer of the organism must be equal to its heat output.

The article considers continuous and discrete-vascular models, analyzes their advantages and disadvantages. Continuous models are, in essence, a simplified notation of biothermal equations. In these models, the effect of blood flow on each individual vessel is neglected, and the blood supply is averaged over the volume studied. Discrete-vascular models are a set of biothermal equations that describe the blood flow in each individual vessel. Discrete-vascular models of the thermophysiological state of man are not very applicable today, due to the complex and insufficiently studied vascular geometry. More common are continuum models, for the construction of which a multi-compartmental approach is used.

As a result, the article presents a comparative table of continuous models and information systems that use these models.

Keywords: *mathematical modeling, heat transfer, information systems, thermophysiological state of man.*



Микола Пирог. Закінчив факультет інформаційних систем і технологій Приватного вищого навчального закладу «Європейський університет», Київ, Україна. Працює асистентом на кафедрі прикладних інформаційних систем факультету інформаційних технологій Київського національного університету імені Тараса Шевченка. Наукові інтереси: інформатизація суспільства, електронне урядування, agile-методології у програмуванні й освіті.

Mykola Pyroh. He graduated from the Faculty of Information Systems and Technologies of the Private Higher Educational Institution "European University". He works as an assistant at the Department of Applied Information Systems of Faculty of Information Technology of Taras Shevchenko National University of Kyiv. Kyiv, Ukraine. Research interests: informatization of society, e-government, agile-methodologies in programming and education.



Ксенія Духновська. Закінчила механіко-математичний факультет Київського національного університету імені Тараса Шевченка. Працює на кафедрі прикладних інформаційних систем факультету інформаційних технологій Київського національного університету імені Тараса Шевченка на посаді асистента, Київ, Україна. Наукові інтереси: медична кібернетика, NLP.

Ksenia Dukhnovska. She graduated from the Faculty of Mechanics and Mathematics of Taras Shevchenko National University of Kyiv. Works at the Department of Applied Information Systems of Faculty of Information Technology of Taras Shevchenko National University of Kyiv as an assistant. Kyiv, Ukraine. Research interests: medical cybernetics, NLP.



Ірина Гарко. Закінчила фізико-математичний факультет Національного педагогічного університету імені М. П. Драгоманова. Працює доцентом кафедри прикладних інформаційних систем факультету інформаційних технологій Київського національного університету імені Тараса Шевченка, канд. фіз.-мат. наук, Київ, Україна. Наукові інтереси: теорія ймовірностей і математична статистика, фрактальний аналіз, системи числення, agile-методології у програмуванні й освіті.

Iryna Harko. Candidate of Physical and Mathematical Sciences. She graduated from the Faculty of Physics and Mathematics of the National Pedagogical University named after M.P. Drahomanov. Works as an associate professor of the Department of Applied Information Systems of Faculty of Information Technology of Taras Shevchenko National University of Kyiv. Kyiv, Ukraine. Research interests: probability theory and mathematical statistics, fractal analysis, number systems, agile methodologies in programming and education.



УДК 004.91

DOI: <https://doi.org/10.17721/AIT.2021.1.05>

Г. М. Гнатієнко, orcid.org/0000-0002-0465-5018,
Київський національний університет імені Тараса Шевченка, Київ, Україна,
О. В. Олецький, orcid.org/0000-0002-0553-5915,
Національний університет “Києво-Могилянська Академія”, Київ, Україна

Порівняння методик ранжування науковців на основі індексів Гірша та PageRank: суперечливі ситуації

Описано результати експериментального порівняння методик оцінювання та ранжування науковців і дослідників на основі їхньої публікаційної активності. Порівнюються наукометричні індикатори на основі індексів Гірша та PageRank-подібних алгоритмів; використовується матриця посилань між статтями, написаними різними авторами. Наведено приклад ситуації, коли застосування різних методик приводить до протилежних результатів – наприклад, коли автори статей, які отримали найкращі оцінки за PageRank, мають найнижчі (тобто найгірші) індекси Гірша. Описана ситуація характеризується тим, що деякий автор отримав ключовий результат, на який посилаються всі інші автори, але самих статей у нього мало, а також із можливими посиланнями в роботі такого автора, статті якого мають високий рівень цитування. Робиться спроба зіставлення результатів із можливими результатами експертного оцінювання, яке має особливе значення за побудови автоматизованих систем, що базуються на комбінуванні різних методик оцінювання. Розглянуто можливе застосування методу аналізу ієрархій. Для побудови матриць попарних порівнянь використано транзитивні шкали з параметром, який задає відношення переваги між градаціями у параметризованому вигляді.

Ключові слова: наукометрія, індекс Гірша, PageRank, метод аналізу ієрархій.

Для цитування (for citation): Г. М. Гнатієнко, О. В. Олецький. “Порівняння методик ранжування науковців на основі індексів Гірша та PageRank: суперечливі ситуації,” *Сучасні інформаційні технології*, vol. 1, p. 39–44, 2021.

ВСТУП

Проблеми оцінювання результативності наукової роботи протягом багатьох століть привертати увагу вчених усього світу. Нині особливо інтенсивно розвиваються методи наукометрії, які використовують, зокрема, для оцінювання та ранжування науковців і дослідників, наукових та освітніх інституцій, наукових журналів відповідно до їхніх здобутків, зокрема і на основі наявних статей. У цьому разі за основу беруть такі показники: кількість статей, кількість посилань на них тощо. На підставі цього для кожного дослідника можна розрахувати його індекс Гірша, або h-індекс [1], який належить до так званих author-level metrics [2].

Індекс Гірша прийнято розглядати як один з основних наукометричних показників для порівняння науковців між собою. Він має позитивні риси, але, як і інші наукометричні показники, містить і суттєві недоліки. Існує ряд його модифікацій, а також пропозицій щодо альтернативних показників. Слід зазначити, що в дослідженні різних методик ранжування та порівняння на

основі наукометричних показників домінують евристичні підходи, у яких суттєву роль при зіставленні та ранжуванні досягнень науковців відіграють суб'єктивні чинники – введені для аналізу евристики. Кожний показник має свої переваги та недоліки, і не можна однозначно сказати, який із цих показників приводить до кращих результатів.

Цікавими є ситуації, коли застосування різних методик приводить до протилежних результатів. У цій роботі наведено один із таких прикладів, отриманий у ході експериментального порівняння методик ранжування на основі індексу Гірша та на основі алгоритму PageRank [3, 4]. Здійснено також зіставлення цих результатів із можливими результатами експертного оцінювання на основі побудови матриць попарних порівнянь.

Метою роботи є проведення дослідження ранжування результативності науковців за їхніми науковими працями та посиланнями на них на основі застосування різних методик оцінювання та зіставлення одержаних результатів.

© Гнатієнко Г. М., Олецький О. В., 2021



ПОСТАНОВКА ЗАДАЧІ

Нехай задано такі величини:

A – множина авторів;

S – множина статей;

$G(a)$ – множина статей, написаних автором a ;

$Q(a) = |G(a)|$ – кількість статей, написаних автором;

$K(s)$ – кількість посилань на статтю s ;

$C(a)$ – кількість посилань на всі статті автора a ;

$H(a)$ – індекс Гірша автора a .

Необхідно дослідити публікаційну активність авторів та їхні показники цитованості, застосовуючи різні підходи. Зокрема, для здійснення цього є сенс застосувати h -індекс (індекс Гірша) й алгоритми PageRank.

ОГЛЯД ДОСЛІДЖЕНЬ ЗА ТЕМАТИКОЮ

У наукових дослідженнях за ступенем складності прийнято розрізняти структуровані, слабкоструктуровані та неструктуровані проблеми. Для підвищення рівня структурованості проблеми застосовують різні підходи та прийоми. Зокрема, практика упорядкування об'єктів, які досліджуються, сприяє структуризації проблеми. Відтак це розкриває можливості для поглибленого вивчення та застосування досконаліших та обгрунтованіших математичних методів.

Однак у задачах, де присутня суб'єктивна складова, завжди існують ситуації, які можуть привести до протилежних результатів. У цій статті розглянуто ситуацію, яка демонструє можливості одержання діаметрально протилежних результатів у разі застосування різних евристик. Прийняття рішень на основі застосування такого виду аналізу має бути ретельно зваженим і всебічно дослідженим.

Тому підвищення рівня структурованості проблеми полягає не тільки в упорядкуванні альтернативи, але й у дослідженні та порівнянні евристик, які застосовують для досягнення цієї мети. Зокрема, перспективним також є використання комплексних підходів з урахуванням зіставлення евристик і врахування їхніх особливостей.

Проблема, яка розглядається у цій статті, є міждисциплінарною. Для її розв'язання слід залучити інструментарій кількох наукових напрямів:

- дослідження наукового простору і, зокрема, наукометричні дослідження;

- експертні технології у цілому і задачі ранжування альтернатив і визначення їхньої ваги, зокрема;

- застосування методів багатокритеріальної оптимізації тощо.

Науковим простором є множина структурно упорядкованих об'єктів, яку можна представити

моделлю, що відображує структуру групування й упорядкування елементів цього простору [5, 6]. Епоха цифровізації, в якій ми живемо, вимагає від наукової спільноти сприйняття ідеї наукометричності та широкого застосування її в науковій діяльності [7, 8].

Досить поширеною думкою є ідея про те, що мірою корисності публікації є її цитованість [9, 10]. На основі цього постулату виокремлюють групу показників, які використовують кількість публікацій науковця та кількість цитувань цих публікацій. Серед наукометричних показників особливе місце належить індексу Гірша. Наприклад, у роботі [11] виділено 12 «гіршоподібних» наукометричних показників.

Слід зазначити, що вивчаючи інформацію, яка стосується наукових досліджень, та аналізуючи її, нині найчастіше застосовують методи математичної статистики. Для підвищення рівня досліджень потрібно розробити відповідні математичні моделі та застосовувати у цій галузі методи теорії прийняття рішень, методи розв'язання задач багатокритеріальної оптимізації, експертні технології, «м'які» обчислення тощо.

МЕТОДИ Й ІНСТРУМЕНТИ ДОСЛІДЖЕННЯ

Зазначимо, що

$$C(a) = \sum_{s \in G(a)} K(s)$$

(принаймні в найпростішому випадку). Можна розглядати не просто кількість посилань, а кількість статей, з яких є посилання на статті автора, але це окреме питання.

Автор a має індекс Гірша r (тобто $H(a) = r$), якщо він має r публікацій, на кожен з яких є не менше ніж r посилань. Причому вважають, що наявність посилань свідчить про достатньо високу якість статті та її важливість для відповідного наукового напрямку. Значну роль відіграє і авторитет автора статті.

Неявно враховується і кількість статей, оскільки очевидно, що

$$\forall a: H(a) \leq Q(a). \quad (1)$$

Проте тут є принципове обмеження, яке виникає у разі досягнення рівності, у яку може перетворитися нерівність (1):

$$H(a) \leq Q(a). \quad (2)$$

Для ситуації, яка виражена рівністю (2), подальше збільшення кількості посилань, хоч воно само по собі є позитивним фактором, на зростання індексу Гірша вже не впливає. Якість відповідних статей, так само як і якість посилань на них, враховується недостатньо.



Алгоритм PageRank [3, 4] був успішно застосований у Google для ранжування вебсторінок. Але його можна використати для ранжування вузлів довольного графа на основі аналізу зв'язків між ними.

Базовий варіант алгоритму можна охарактеризувати так.

Нехай є n вузлів (вершин графа); граф вважаємо орієнтованим. Нехай граф задано матрицею суміжностей M , елементи якої отримують за таким принципом:

$$M_{ij} = \begin{cases} 1, & \text{якщо } i\text{-й вузол посилається на } j\text{-й,} \\ 0 & \text{– в іншому випадку,} \end{cases} \quad (3)$$

$i, j = 1, \dots, n.$

На основі матриці M з елементами вигляду (3) будується матриця перехідних імовірностей P (тут ми дотримуємося стандартної імовірнісної термінології, пов'язаної з випадковим блуканням і відповідним марковським ланцюгом, хоча з тим же успіхом можна розглядати і неймовірнісні інтерпретації):

$$P_{ij} = (1 - \alpha) / n + \alpha \times r_{ij},$$

$$r_{ij} = M_{ij} / s_i, \quad i, j = 1, \dots, n,$$

де s_i – кількість вихідних дуг з i -ї вершини;

α – деяке задане значення;

$0 < \alpha < 1$ – параметр алгоритму; зокрема при ранжуванні веб-сторінок часто приймається $\alpha = 0,85$.

Можна сказати, що сформульований таким чином класичний алгоритм PageRank враховує не тільки кількість посилань на вузол, але й вагу цих посилань: посилання від вузлів із більш високим рангом мають більшу вагу.

У нашій задачі вузлами графа виступають статті (це відмінність, зокрема, від роботи [2], в якій вузли відповідають авторам). Відповідно,

$M_{ij} = 1$, якщо стаття s_i посилається на статтю s_j .

Тоді, очевидно, що

$$K(s_j) = \sum_{i=1}^n M_{ij}. \quad (4)$$

Описана методика призначена для ранжування статей на основі значень (4). Проте зрозуміло, що її з очевидними модифікаціями можна використати і для оцінювання та ранжування авторів цих статей. Так, оцінка автора $V(a)$ може бути отримана на основі оцінок його статей $v(s_1), v(s_2), \dots$ наприклад, за однією з формул:

$$V(a) = \max_{s \in G(a)} v(s) \quad (5)$$

(оцінка найкращої статті автора) або

$$V(a) = \sum_{s \in G(a)} v(s) / |G(a)|. \quad (6)$$

(середня оцінка за всіма статтями автора).

Крім формул (5) та (6) можна використати і комбіновану оцінку автора на основі обчислення зваженої суми

$$V(a) = \sum_{s \in G(a)} \rho(s) v(s), \quad (7)$$

тобто, коли різним статтям авторів будуть приписуватися різні ваги $\rho(s)$ залежно від їхніх рангів. Утім, приклад, наведений у статті, носитиме досить базовий характер, і для цих цілей конкретна формула вигляду (5), (6) чи (7) не матиме особливого значення.

ПОРІВНЯННЯ ПІДХОДІВ ТА ЕКСПЕРТНЕ ОЦІНЮВАННЯ

Розглянемо конкретний приклад, який ілюструє описані вище підходи. Цей приклад показує ситуацію суперечності, коли автор статті, найкращої за критерієм PageRank, отримує найменший індекс Гірша.

Нехай перший автор є автором ключової, тобто найважливішої статті, результати якої використовують інші автори, і відповідно на яку є посилання з усіх інших статей. Проте він є автором лише однієї статті, і тому його індекс Гірша не може бути більшим за 1. Статті ж інших авторів мають посилання одна на іншу.

У цьому прикладі є 4 автори, 9 статей. Статті розподілено між авторами таким чином (тут нумерація починається з 0):

$$G(0) = \{0\}, G(1) = \{1,8\}, G(2) = \{2,3,4\},$$

$$G(3) = \{5,6,7\}.$$

Отже,

$$Q(0) = 1, Q(1) = 2, Q(2) = Q(3) = 3.$$

Граф, який задає структуру посилань з однієї статті на іншу, задано такою матрицею суміжності:

$$M = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Для параметра α візьмемо одне з типових значень: $\alpha = 0,85$.

Міри важливості вузлів (статей), що розраховані за алгоритмом PageRank, мають вигляд:



0,3431 0,3207 0,0509 0,0509 0,0509
0,0518 0,0518 0,0509 0,0292

Таким чином, 0-ва стаття (вона ж єдина стаття 0-го автора) отримує перевагу за PageRank, як і слід було чекати. На цьому ж прикладі можна подивитися підвищення ваги наступної статті за рахунок того, що на неї посилається найкраща. Хоча слід зазначити, що в цьому випадку таке підвищення може не розглядатися як особливо заслужене (не виключена ситуація, що автор цієї статті міг якось домовитися з автором найкращої статті).

На цій основі можна розрахувати оцінки авторів. Так, якщо оцінку автора обчислювати як максимум оцінок окремих статей, то маємо:

$$V(0) = 0,3431; V(1) = 0,3207; V(2) = 0,0059; \\ V(3) = 0,0518.$$

Проте значення індексів Гірша авторів такі:

$$H(0) = 1, H(1) = 2, H(2) = 3, H(3) = 3.$$

Таким чином, експеримент показує, що автори двох статей, які отримали найвищі (тобто найкращі) оцінки за PageRank, мають найнижчі (тобто найгірші) індекси Гірша.

У подібних ситуаціях часто доцільно залучати до розгляду методики експертного оцінювання [12, 18 та ін.]. Зіставлення результатів роботи різних алгоритмів може виявитися корисним у ситуаціях, коли доводиться робити висновки, які методики автоматизованого алгоритмічного оцінювання заслуговують на більшу довіру або, можливо, навпаки.

Особливу увагу в цьому контексті слід звернути на метод попарних порівнянь, який часто використовують у межах методу аналізу ієрархій Сааті [13, 14 та ін.].

У прикладі, що розглядається, видаються обґрунтованими такі міркування експертів. Імовірно, що вони надали б перевагу першому (0-му) автору, оскільки на його роботу спираються всі інші автори. Наступні два автори могли б розглядатися як рівнозначні. Останній мав би отримати найгіршу оцінку – оскільки на його роботи немає жодних посилань, крім самоцитування та сумнівного посилання від 0-го автора.

Матриця попарних порівнянь у такій ситуації могла б мати такий вигляд:

$$\begin{pmatrix} 1 & 2 & 2 & 3 \\ 1/2 & 1 & 1 & 2 \\ 1/2 & 1 & 1 & 2 \\ 1/3 & 1/2 & 1/2 & 1 \end{pmatrix}.$$

Для елементів цієї матриці, відповідно до стандартного підходу, значення 2 означає мінімальну перевагу одного варіанта над іншим, 3 – дещо більш виражену перевагу). Деякі автори звертали увагу на те, що застосування саме такої шкали градацій не завжди приводить до добрих результатів, зокрема через те, що розкид між максимальною та мінімальною оцінкою може

вийти надто великим, і в таких випадках може виявитися доцільним застосування так званих транзитивних шкал [14–16 та ін.]. Для транзитивних шкал вводиться деякий параметр τ , який визначає, наскільки наступний рівень градацій порівняння є більшим за попередній. Наприклад, якщо рівнозначність альтернатив оцінюється значенням 1, а мінімальний рівень переваги оцінюється як τ , то наступна градація переваг буде оцінена як τ^2 , і т. д. Але це питання заслуговує на окремий розгляд.

Зазначимо, що для оцінювання переваг між варіантами може бути застосована шкала Томаса Сааті з градаціями (1, 3, 5, 7, 9) або шкала оцінювання задач, яка використовується у методології Agile з градаціями (1, 2, 3, 5, 8, 13, 21). Проте дослідження застосування різних типів шкал не є предметом цієї статті і може бути здійснене у подальшому в окремій роботі.

Для наведеної матриці попарних порівнянь її нормалізований головний власний вектор, кожна компонента якого є оцінкою відповідного автора, дорівнює

$$0,4236 \quad 0,2270 \quad 0,2270 \quad 0,1223.$$

Таким чином, 0-й автор перемагає, із цього погляду оцінки на основі PageRank виявляються ближчими до результатів експертного оцінювання порівняно з оцінками на основі аналізу індексів Гірша.

Проілюструємо також можливе використання параметризованих транзитивних шкал. Параметризована матриця попарних порівнянь для нашого прикладу могла б мати вигляд

$$\begin{pmatrix} 1 & \tau & \tau & \tau^2 \\ \frac{1}{\tau} & 1 & 1 & \tau \\ \frac{1}{\tau} & 1 & 1 & \tau \\ \frac{1}{\tau^2} & \frac{1}{\tau} & \frac{1}{\tau} & 1 \end{pmatrix}.$$

При $\tau = 1.1$ вектор оцінок складатиметься з елементів, які мають такі значення:

$$0,2744 \quad 0,2494 \quad 0,2494 \quad 0,2268.$$



Таким чином, зі зменшенням τ розкид між оцінками суттєво зменшився. Вибір параметра τ може мати суттєве значення для застосування алгоритмів оцінювання. Деякою мірою питання про вибір τ обговорювалося в [16] для задачі автоматизованого оцінювання студентських учбових проєктів. Проте, узагалі кажучи, питання застосування транзитивних шкал і вибору параметра τ , зокрема і для аналізу наукометричних даних, заслуговує на окремих розгляд.

ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

Для детальнішого зіставлення евристик, які застосовуються у наукометричних дослідженнях, слід додатково дослідити різні підходи та зіставити результати не лише у двох описаних напрямках, але порівняти ширший набір інструментів, які застосовують дослідники.

Слід також запропонувати критерії вибору інструментальних засобів для наукометричних досліджень, виявити позитивні та негативні риси використовуваних евристик.

ВИСНОВКИ

Застосування єдиного критерію у наукометричних дослідженнях є зручним інструментом, але часто – нерепрезентативним та одностороннім. Цей інструмент у багатьох практичних ситуаціях не відображає належною мірою переможців серед науковців, і лише висвітлює одну грань наукових досягнень.

Наведений приклад експериментально підтверджує, що ранжування авторів на основі аналізу індексів Гірша може суттєво відрізнятись від ранжувань, отриманих на основі оцінок PageRank для статей, і на основі цього – відрізнятись для авторів відповідних статей. Справді, експеримент для наведеного прикладу показує, що автори статей, найкращих за PageRank, мають найгірші індекси Гірша.

Проте слід зазначити, що ранжування як на основі індексів Гірша, так і на основі PageRank, має свої переваги та недоліки – як, власне, й інші методики. Тому може виявитись доцільним комбінування різних методик – зокрема, шляхом аналізу того, яким чином методики автоматизованого алгоритмічного оцінювання забезпечують дотримання певних принципів і політик оцінювання. Можна, наприклад, говорити про надання окремих методикам певних коефіцієнтів надійності, і про налаштування цих коефіцієнтів на основі застосування тих чи інших процедур машинного навчання, зокрема і навчання з підкріпленням [17, 19].

Можна зробити однозначний висновок, що дослідження результативності науковців має бути різностороннім, а застосування інструментів – комплексним. Для порівняння й адекватного комбінування різних методик може виявитись дуже корисним застосування методів експертного оцінювання; у статті наводиться відповідний приклад. У цьому контексті варто також звернути увагу на роботу [20], в якій ідеться про аналіз різних методик оцінювання дослідників, зокрема на основі використання наукометричних показників.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

- [1] J.E. Hirsch, “An index to quantify an individual's scientific research output”, *PNAS*, vol.102 (46), pp.16569–16572, 2005.
- [2] J.D. West, M.C. Jensen, R.C. Dandrea, G.J. Gordon, C.T. Bergstrom. “Author-level Eigenfactor metrics: Evaluating the influence of authors, institutions, and countries within the social science research network community”, *Journal of the American Society for Information Science and Technology*, vol.64(4), pp.787–801, 2013.
- [3] S. Brin, L. Page. “The anatomy of a large-scale hypertextual web search engine”, in *Proc. WWW*, 1998, pp.107–117.
- [4] К.Д. Маннинг, П. Рагхаван, Х. Шютце, *Введение в информационный поиск*, Москва: ООО «И.Д. Вильямс», 2011.
- [5] В.Г. Кремень, В.Ю. Биков, “Категорії «простір» і «середовище»: особливості модельного подання та освітнього застосування”, *Теорія і практика управління соціальними системами*, no.2, pp.3–16, 2013.
- [6] О.Ю. Кучанський, “Складові аналізу наукових мереж”, *Управління розвитком складних систем*, no.41, pp.115–126, 2020.
- [7] А.В. Кільченко, “Вітчизняний та зарубіжний досвід використанні інформаційно-цифрових технологій для оцінювання результативності науково-педагогічних досліджень”, in *Звітна науково-практична конференція Інституту інформаційних технологій і засобів навчання НАПН України : матеріали науково-практичної конференції*, 2021, pp.48–54.
- [8] В.Ю. Биков, О.М. Спірін, А.О. Білощицький та ін., “Відкриті цифрові системи в оцінюванні результатів науково-педагогічних досліджень”, *Інформаційні технології і засоби навчання*, vol.75(1), pp.294–315, 2020.
- [9] В.М. Горючий, “Критерії якості наукових досліджень у контексті забезпечення національних інтересів”, *Вісник Національної академії наук України*, no.6, pp.74–80, 2015.
- [10] О.М. Спірін, С.М. Іванова, А.В. Кільченко, Т.Д. Новицька, “Використання наукометричних баз даних і систем вебаналітики для моніторингу електронних наукових фахових видань”, *Інформаційні технології в освіті*, vol.45(4), pp.18–30, 2020.
- [11] С.Д. Штовба, Е.В. Штовба, “Обзор наукометрических показателей для оценки публикационной деятельности ученого”, *Управление большими системами. Спец. вып. 44: Наукометрия и экспертиза в управлении наукой*, pp.262–278, 2013.
- [12] В.Є. Снитюк, Г.М. Гнатієнко, *Експертні технології прийняття рішень*, Київ: Маклаут, 2008.



[13] Т. Саати, *Принятие решений. Метод анализа иерархий*, Москва: Радио и связь, 1993.

[14] И. Г. Черноруцкий. *Методы принятия решений*, Санкт-Петербург: БХВ-Петербург, 2005.

[15] O. V. Oletsy, E.V. Ivohin, "Formalizing the Procedure for the Formation of a Dynamic Equilibrium of Alternatives in a Multi-Agent Environment in Decision-Making by Majority of Votes", *Cybern Syst Anal*, vol.57, pp.47-56, 2021.

[16] А. В. Олецкий, М.Ф. Махно, "О повышении уровня адекватности в результатах процесса оценивания учебных проектов на основе параметрической релаксации метода парных сравнений", *Проблемы управления и информатики*, no.1, pp.122–133, 2021.

[17] С. Рассел., П.Норвиг, *Искусственный интеллект: современный подход*, Москва: Изд. дом «Вильямс», 2006.

[18] M. Antonewich, A. Didyk, and V. Snytyuk, "Optimization of Functions of Two Variables by Deformed Stars Method," in *Proc. ATIT*, 2019, pp. 475–480.

[19] С. М. Николенко, А.Л. Тулупьев. *Самообучающиеся системы*, Москва: МЦНМО, 2009.

[20] Н. Hnatiienko, V. Snytyuk, N. Tmienova, O. Voloshyn, "Determining the effectiveness of scientific research of universities staff" in Selected Papers of the 7th International conference "Information Technology and Interactions" (IT&I-2020). *CEUR Workshop Proceedings*, 2883, 2021, pp. 164–176.

[21] M. Shell. (2002) IEEEtran homepage on CTAN. [Online]. Available: <http://www.ctan.org/tex-archive/macros/latex/contrib/supported/IEEEtran/>

Стаття надійшла до редколегії

03.10.2021

Comparing methods of ranking scientists on the base of Hirsch index and PageRank: contradictory situations

Experiments aimed at comparing different methods of estimating and ranking scientists and researchers on the base of their publication activity are reported. Scientometric indicators based on h-index and PageRank are being compared. For such a comparison, a graph of citations represented by a matrix was applied. An example when different methods lead to opposite results was described. For example, authors having the best PageRank-based estimations may have the least h-indices. Such a situation is possible when a high-cited author managed to obtain a key result cited by all the other authors but this author has few papers. A comparison with methods of expert estimations was carried out, which appears to be very useful for building automated systems combining various methods of algorithmic estimating and ranking. The Analytic Hierarchy Process was applied. For building pairwise comparison matrices, transitive scales with a parameter representing how much times the next level of advantage is bigger than the previous one were harnessed.

Keywords: scientometrics, h-index, PageRank, Analytic Hierarchy Process.



Григорій Гнатієнко. Закінчив факультет кібернетики Київського національного університету імені Тараса Шевченка, канд. техн. наук. Місце роботи: факультет інформаційних технологій Київського національного університету імені Тараса Шевченка, Київ, Україна. Посада: заступник декана з наукової роботи. Наукові інтереси: оброблення експертної інформації, прийняття рішень, інформаційні технології.

Hryhorii Hnatiienko. PhD. Education: Faculty of Cybernetics, Taras Shevchenko National University of Kyiv. Affiliation: Faculty of Information Technologies of Taras Shevchenko National University of Kyiv. Kyiv, Ukraine. Position: Deputy Dean for Research. Research interests: processing of expert information, decision making, information technologies.



Олексій Олецкий. Закінчив факультет кібернетики Київського національного університету імені Тараса Шевченка, канд. техн. наук. Місце роботи: факультет інформатики, Національний університет «Києво-Могилянська академія», Київ, Україна. Посада: доцент кафедри мультимедійних систем. Наукові інтереси: прийняття рішень, штучний інтелект.

Oleksiy Oletsy. PhD, Associate Professor. Education: Faculty of Cybernetics, Taras Shevchenko National University of Kyiv. Affiliation: Faculty of Informatics, National University of Kyiv-Mohyla Academy. Kyiv, Ukraine. Position: Associate Professor. Research interests – decision making, artificial intelligence.



МАШИННЕ НАВЧАННЯ ТА РОЗПІЗНАВАННЯ ОБРАЗІВ



УДК 004.934

DOI: <https://doi.org/10.17721/AIT.2021.1.06>О. Є. Іларіонов, orcid.org/0000-0002-7435-3533,

А. К. Астахов,

Г. В. Красовська, orcid.org/0000-0003-1986-6130,І. М. Доманецька, orcid.org/0000-0002-8629-9933,

Київський національний університет імені Тараса Шевченка, Київ, Україна

Інтелектуальний модуль розпізнавання емоцій за голосом

Для людей мовлення є основним засобом комунікації, причому люди з мовлення можуть отримувати не тільки семантичну, а й емоційну інформацію. Розпізнавання емоцій за голосом є актуальним для таких галузей, як надання психологічної допомоги, розроблення систем безпеки, виявлення брехні, аналіз зв'язків із клієнтами, розроблення відеоігор. Оскільки розпізнавання емоцій людиною є суб'єктивним, а отже неточним, та потребує багато часу, існує необхідність у створенні програмного забезпечення, яке могло б розв'язати цю задачу. У статті розглянуто стан проблеми розпізнавання емоцій людини за голосом. Проаналізовано сучасні публікації, використані в них підходи, а саме: моделі емоцій, набори даних, методи вилучення ознак, класифікатори. Визначено, що існуючі розробки мають середню точність близько 0,75. Проаналізовано загальну структуру системи розпізнавання емоцій людини за голосом, спроектовано та розроблено відповідний інтелектуальний модуль. За допомогою уніфікованої мови моделювання UML (від англ. "Unified Modeling Language") створено діаграму компонентів і діаграму класів. Як набори даних обрано датасети RAVDESS і TESS для урізноманітнення навчальної вибірки. Використано дискретну модель емоцій (радість, смуток, гнів, відраза, страх, здивування, спокій, нейтральна емоція), метод MFCC (мел-частотні кепстральні коефіцієнти від англ. "Mel Frequency Spectral Coefficients") для вилучення ознак, згорткову нейронну мережу для класифікації. Нейронну мережу розроблено з використанням бібліотек для машинного навчання TensorFlow і Keras. Побудовано спектральну графіку аудіосигналу, а також графіку точності та похибки розпізнавання. За результатами програмної реалізації інтелектуального модуля розпізнавання емоцій за голосом збільшено точність валідації до 0,8.

Ключові слова: розпізнавання емоцій за голосом, нейронні мережі, глибинне навчання, згорткові нейронні мережі.

Для цитування (for citation): О. Іларіонов, А. Астахов, Г. Красовська, І. Доманецька "Інтелектуальний модуль розпізнавання емоцій за голосом", *Сучасні інформаційні технології*, vol.1, p. 46–52, 2021.

ВСТУП

Для людей мова є найприроднішим способом самовираження та передачі інформації. Найкраще ми усвідомлюємо її важливість, коли нам доводиться використовувати інші способи спілкування, такі як електронна пошта чи текстові повідомлення. У цьому разі нам одразу стає важче зрозуміти співрозмовника, адже ми звикли отримувати не тільки семантичну інформацію зі слів, а й мати справу з емоціями людини, які допомагають нам краще зрозуміти ситуацію. Тому логічним є поширення цього розуміння і на комп'ютери. Розпізнавання мови вже присутнє у нашому повсякденному житті завдяки розумним мобільним пристроям, які здатні приймати голосові команди та відповідати

на них синтезованою мовою. Задача розпізнавання емоцій за голосом існує вже більше двох десятиліть [1], є програми для взаємодії людина-комп'ютер [2], а також роботи [3], мобільні послуги [4], колцентри [5], комп'ютерні ігри [6] та психологічна оцінка [7, 8]. Ця задача є вкрай складною через те, що емоції суб'єктивні. Не існує єдиного консенсусу щодо того, як їх вимірювати або класифікувати.

МЕТА ДОСЛІДЖЕННЯ

Зрозуміло, що вже є публікації, в яких описано існуючі дослідження з розпізнавання емоцій за голосом. Були розглянуті в основному ті дослідження, що опубліковані порівняно недавно, адже попередні публікації не могли

© Іларіонов О. Є., Астахов А. К., Красовська Г. В., Доманецька І. М., 2021



включати останні досягнення та тенденції, такі як нейронні мережі.

У своєму дослідженні щодо розпізнавання емоцій за голосом (2006) автори [9] спеціально зосередилися на збиранні мовних даних, виокремлюючи акустичні особливості для задачі класифікації. У сучаснішій роботі [10] представлено дослідження, присвячене аналізу особливостей мовлення з використанням відеоінформації. Коолагуді та Рао також займалися класифікацією баз даних, методів вилучення ознак і побудови класифікаторів для своїх досліджень [11].

Анагностопулос та Джаннукос провели всебічний аналіз публікацій 2000–2011 рр. [12]. Їхнє дослідження – одне з перших, що стосується глибинних нейронних мереж у задачі розпізнавання емоцій за голосом. Вони також висвітлюють дослідження, які використовують гібридні класифікатори, та схеми голосування.

Дослідження Рамакрішнана включає не лише бази даних, методи вилучення ознак і класифікатори, але також розглядає нормалізацію сигналів, що є етапом передоброблення, яка обов'язково виконується перед вилученням ознак [13]. У цій роботі запропоновано нові області застосування систем розпізнавання емоцій за голосом, що впливає на вимоги та методи розроблення.

Нещодавно проведено короткий аналіз баз даних, методів зменшення шуму для попереднього оброблення сигналів, методів вилучення ознак і класифікаторів, включаючи останні досягнення, такі як згорткові та рекурентні нейронні мережі [14].

Зазначимо, що нещодавно опубліковано роботу, в якій проаналізовано розпізнавання емоцій за текстом і голосом, причому в ній детально розглянуто різні моделі емоційних

станів. У дослідженні [15] виконано аналіз різних класифікаторів.

З урахуванням проаналізованих робіт можна дійти таких висновків:

- основними моделями емоцій є багатовимірні та дискретні моделі;
- найуживанішими методами вилучення ознак є спектральні, а саме MFCC (мел-частотні кепстральні коефіцієнти);
- середня точність класифікаторів складає близько 0,75; причому найточнішими виявилися класифікатори, побудовані на базі прихованих марковських моделей і нейронних мереж; найперспективнішими класифікаторами вважають саме нейронні мережі.

Метою статті є аналіз сучасного стану проблеми розпізнавання емоцій за голосом, дослідження емоційних моделей, методів вилучення ознак, наборів даних і класифікаторів, розроблення та тестування інтелектуального модуля розпізнавання емоцій людини за голосом.

ВИКЛАД ОСНОВНОГО МАТЕРІАЛУ

Система розпізнавання емоцій за голосом — це сукупність методологій, які обробляють та класифікують мовні сигнали для виявлення в них емоцій. Загалом, система має такі складові:

- 1) модель емоцій;
- 2) набір даних;
- 3) процеси передоброблення;
- 4) вилучення ознак;
- 5) класифікація.

На рис. 1 зображено структурні елементи системи та технологія розв'язання задачі розпізнавання емоцій за голосом. На вхід система приймає аудіозапис мовлення, далі відбувається оброблення сигналу, накладання моделі емоцій, виокремлення ознак і розпізнавання емоцій.

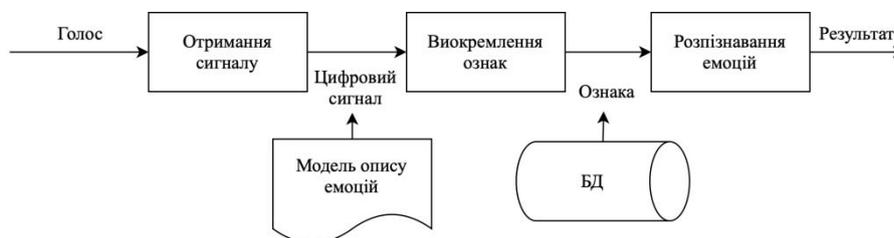


Рис. 1. Система розпізнавання емоцій за голосом

Для успішної реалізації системи розпізнавання емоцій за голосом потрібно правильно визначати та моделювати емоції. Однак єдиного консенсусу щодо визначення емоції не існує, і це все ще залишається відкритою проблемою у психології. У ХХ ст. запропоновано понад

90 визначень емоцій. Загалом, можна сказати, що емоції є деякими психологічними станами, що складаються з декількох компонентів, таких як особистий досвід, фізіологічні, поведінкові та комунікативні реакції. На основі цих визначень у розпізнаванні емоцій набули поширення дві



моделі: дискретна модель емоцій і багатовимірна модель емоцій.

Дискретна модель емоцій базується на шести категоріях основних емоцій: смуток, радість, страх, гнів, відраза та здивування. Ці вроджені та незалежні від культури емоції переживаються протягом короткого періоду часу. Інші емоції отримують поєднанням основних. Більшість існуючих систем розпізнавання емоцій зосереджено на вказаних основних емоційних категоріях. Однак цими категоріями емоційних станів не можна вичерпно описати деякі складні емоційні стани.

Багатовимірна модель емоцій – це альтернативна модель, яка використовує невелику кількість вимірів, наприклад, привабливість, збудження, контроль, влада. Однією з найуживаніших моделей є двовимірна модель, яка використовує збудження чи активацію як один вимір, а валентність, чи оцінку – як інший. Валентність визначає, чи є емоція позитивною або негативною, і змінюється від неприємної до приємної. Збудження визначає силу почуттів і може коливатися від нудьги до шаленого хвилювання. Тривимірна модель включає вимір домінування або влади, що характеризує сили людини, і міститься в діапазоні між слабкістю і силою. Наприклад, за цим виміром можна відрізнити гнів від страху. Існує декілька недоліків багатовимірного зображення. По-перше, вказана модель недостатньо інтуїтивна, і для позначення кожної емоції може знадобитися спеціальна підготовка. Крім того, деякі емоції, наприклад, здивування, важко віднести до певної категорії, адже ці емоції можуть мати позитивну чи негативну валентність залежно від контексту.

З урахуванням простоти в роботі та частоти використання під час розв'язання задачі розпізнавання емоцій за голосом, як модель емоційних станів, обрано дискретну модель.

Набори даних для розпізнавання емоцій за голосом можна розділити на три категорії:

- 1) датасети, що складаються з емоцій, які відіграні акторами;
- 2) датасети із записами емоцій, що були штучно викликані;
- 3) датасети із записами природних емоцій.

Перша категорія наборів даних складається з емоцій, що відтворені професійними акторами у звуконепроникних студіях. Створити такий набір даних порівняно простіше, однак цей спосіб може неадекватно передавати силу емоцій, часто призводячи до перебільшення. Це знижує точність розпізнавання у реальному житті.

Для створення другої категорії наборів даних дослідники відтворюють ситуації, що штучно

провокують емоції. Такий спосіб точніше відтворює емоційний стан людини.

Набори даних із записами природних емоцій переважно отримують із токшоу, записів колцентрів, радіопереговорів і подібних джерел. Отримати такі дані важче, у першу чергу через етичні та юридичні проблеми.

Після вибору методу створення набору даних розглядають інші питання проектування, наприклад, вік і стать. Більшість баз даних містять дорослих ораторів, але також є бази даних дітей і людей старшого віку.

Як набори даних пропонується вибрати датасети RAVDESS[16] та TESS[17]. Вони безкоштовні для наукового використання, записані англійською мовою, мають розгорнуту документацію та варіативність у даних.

Використання двох датасетів має на меті покращити процес навчання нейронної мережі, уникаючи перенавчання. Два набори даних містять аудіозаписи з різними параметрами (наприклад, у датасеті RAVDESS вік акторів 21–33 роки, а в датасеті TESS – 26–64 роки). Параметр віку є вкрай важливим для систем розпізнавання емоцій, тому що вираження тієї чи іншої емоції під час мовлення спричинено фізичними особливостями будови голосового тракту людини. Через це з віком змінюється як сам голос, так і патерни, що дозволяють визначати емоції. Тому використання двох датасетів дозволить урізноманітнити навчальну вибірку та збільшити точність класифікації.

Процес передоброблення складається з фреймінгу, віконного перетворення, нормалізації, зменшення шуму.

Фреймінг, також відомий як сегментація мови, виконується для поділу неперервної мови на сегменти фіксованої довжини. Емоції можуть змінюватися у процесі мовлення, однак залишаються незмінними протягом короткого проміжку часу. Цим зумовлена довжина сегментів у 20–30 мс.

Далі виконується віконне перетворення Фур'є. Для цього користуються методом зважування вікном Хеммінга:

$$w(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{M-1}\right),$$

$$0 \leq n \leq M - 1,$$

де M – розмір вікна $w(n)$.

Для етапу нормалізації використовують метод z -нормалізації. При середньому значенні μ та стандартному відхиленні σ , нормалізація розраховується за формулою

$$z = \frac{x - \mu}{\sigma}.$$



Вилучення ознак є вкрай важливим процесом, адже таким чином вдається перетворити аудіофайли у формат, зрозумілий для моделей. Як ознаки використовують мел-частотні кепстральні коефіцієнти (MFCC). Для отримання MFCC висловлювання ділять на сегменти, потім кожний сегмент перетворюють на частотну область за допомогою дискретного перетворення Фур'є. Далі розраховують значення логарифма кожної області, після чого виконують дискретне косинусне перетворення. Коефіцієнтами MFC будуть амплітуди результуючого спектра. Блок-схему отримання кепстральних коефіцієнтів з аудіосигналу зображено на рис. 2.



Рис. 2. Узагальнена блок-схема отримання кепстральних коефіцієнтів

На рис. 3 представлено діаграму компонентів, що описує особливості фізичного подання системи.

Вона складається з компонентів наборів даних (RAVDESS і TESS), двох компонентів сторонніх бібліотек (TensorFlow та Keras), основного класифікатора та інтерфейсу вводу/виводу для роботи з класифікатором.

На рис. 4 зображено діаграму класів. Вона деталізує центральний компонент попередньої діаграми, а саме, модуль розпізнавання емоцій за голосом, і включає такі класи: Datasets (виконує

оброблення датасетів), Features (виконує вилучення ознак), CNN (будує згорткову нейронну мережу), Testing (виконує тестування мережі), Plotting (будує графіки, потрібні для візуалізації). Крім того, на діаграмі наявні два пакети (TensorFlow та Keras), що допомагають у побудові нейронної мережі, та два файли (RAVDESS і TESS), що представляють набори даних.

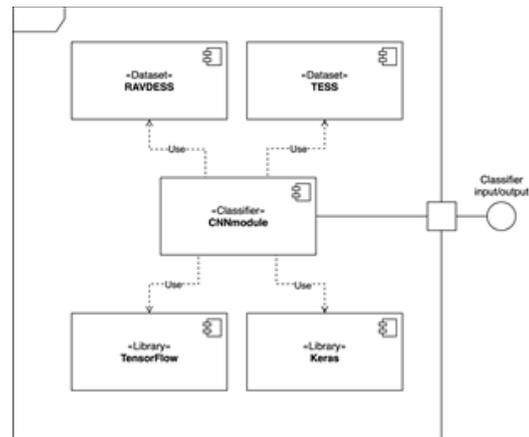


Рис. 3. Діаграма компонентів

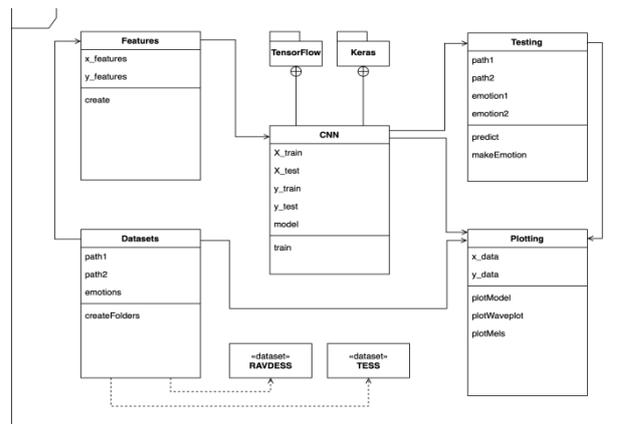


Рис. 4. Діаграма класів

На рис. 5 показано архітектуру нейронної мережі. У першу чергу йдуть основний та обов'язковий шари згортки. Як функцію активації використовують ReLU. Оскільки повнозв'язні шари використовують багато параметрів, вони мають схильність до перенавчання (overfitting). Одним із методів, який дозволяє попереджати перенавчання, є використання так званого дропауту (від англ. *dropout*). На кожному етапі навчання окремі нейрони вилучаються з мережі з деякою ймовірністю $1 - p$. Зв'язки, що вели до нейрона та з нейрона, також вилучаються. Нейрони, які залишилися у мережі, беруть участь у подальшому навчанні. Потім вилучені нейрони повертають до мережі з початковими вагами. Далі додається шар Flatten, метою якого є зведення



даних до одновимірного вектора для того, щоб їх можна було передати на наступний шар. Наступним шаром є повнозв'язний шар, як у звичайного перцептрона. Це стандартний останній шар, який на виході має кількість нейронів, рівну кількості класів. Як функції активації часто використовують Softmax.

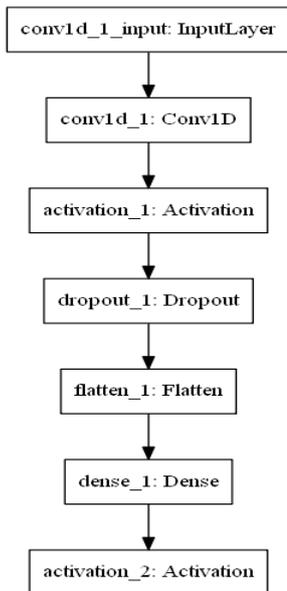


Рис. 5. Архітектура згорткової нейронної мережі

На вхід моделі подається аудіофайл. Мовний сигнал кодується і передається у вигляді акустичного сигналу (звукової хвилі). Основні акустичні параметри мовного сигналу такі: основна частота коливань голосових зв'язок (частота фонації), формантна структура (розташування, амплітуда і ширина формантних областей), рівень звукового тиску, динаміка зміни частоти основного тону (тобто мелодика мови), динамічний діапазон звуку (чинить істотний вплив на тембр голосу), темпо-ритмічні особливості мови, енергетичні характеристики звукового сигналу і їхня зміна в часі та ін.

На рис. 6 зображено приклад графіка аудіо-запису, зробленого чоловіком із нейтральною емоцією.

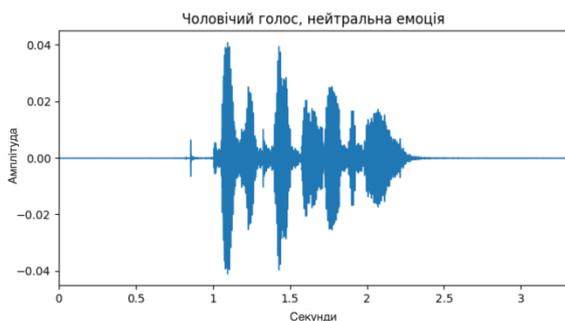


Рис. 6. Графік сигналу нейтральної емоції

На рис. 7 показано приклад Log-Mel-спектрограми для того ж аудіозапису.

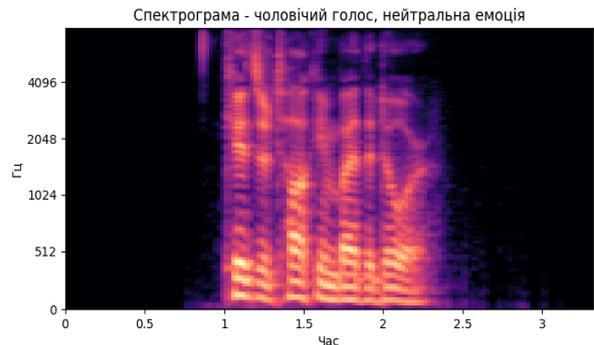


Рис. 7. Спектрограма нейтральної емоції

Мел-кепстральні коефіцієнти мають підвищену завадостійкість і дозволяють приймати достовірні рішення на відносно коротких інтервалах аналізу мовлення. Ознаки, побудовані на основі мел-кепстральних коефіцієнтів, враховують психо-акустичні принципи сприйняття мовлення, оскільки використовують мел-шкалу, пов'язану з так званими критичними смугами слуху.

РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Розроблений інтелектуальний модуль розпізнавання емоцій за голосом показав прийнятний результат роботи.

З набору даних 67% використовували для навчання моделі і 33% для її тестування.

Навчання виконувалося протягом 50 епох.

Графік точності зображено на рис. 8.

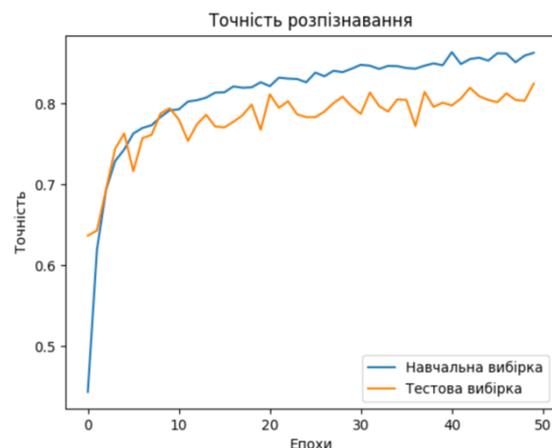


Рис. 8. Графік точності розпізнавання

Точність валідації склала 0.8, що є кращим показником за середній показник точності в проаналізованих публікаціях (0.75).

Ще одним параметром, що характеризує якість навчання мережі є рівень похибки, що демонструє модель у розрізі часу.



На рис. 9 показано графік похибки мережі. Ознаки перенавчання нейронної мережі відсутні.

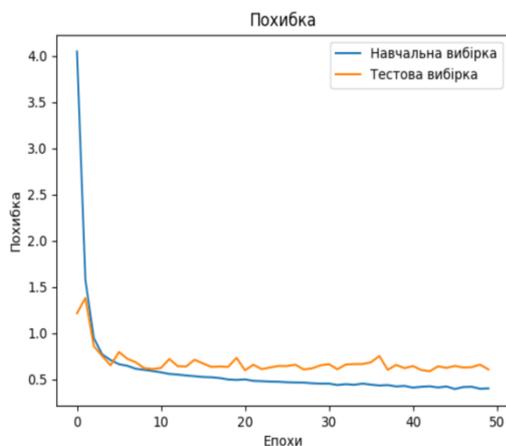


Рис. 9. Графік похибки мережі

Перевірялась точність розпізнавання різних емоцій (гнів, радість, нейтральна, смуток, спокій, страх, відроза, здивування). У таблиці наведено отримані результати, а саме, точність розпізнавання кожного з 8 емоційних станів.

ТАБЛИЦЯ

ТОЧНІСТЬ РОЗПІЗНАВАННЯ ОКРЕМИХ ЕМОЦІЙ

Емоція	Точність
Гнів	0,93
Радість	0,92
Нейтральна	0,91
Смуток	0,84
Спокій	0,96
Страх	0,92
Відроза	0,95
Здивування	0,90

ВИСНОВКИ

Розроблений інтелектуальний модуль розпізнавання емоцій людини за голосом виявився доволі ефективним. Поєднання згорткових нейронних мереж, мел-частотних кепстральних коефіцієнтів і наборів даних RAVDESS і TESS показало точність валідації, вищу за класичні класифікатори.

Планується подальше проведення експериментальних досліджень із використанням інших моделей емоційних станів і наборів даних. Розроблений інтелектуальний модуль можна використати в аналізі роботи колцентрів, у разі психологічного оцінювання, а також для оптимізації роботи інтелектуальних голосових помічників.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

- [1] Schuller, B.W. (2018) «Speech emotion recognition: two decades in a nutshell, benchmarks, and ongoing trends», *Commun. ACM* 61 (5), pp. 90–99. [Online]. Available: doi:10.1145/3129340.
- [2] Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., Taylor, J.G. (2001) «Emotion recognition in human-computer interaction», *IEEE Signal Process. Mag.* 18 (1), pp. 32–80. [Online]. Available: doi:10.1109/79.911197.
- [3] Huahu, X., Jue, G., Jian, Y. «Application of speech emotion recognition in intelligent household robot», in *International Conference on Artificial Intelligence and Computational Intelligence*, 2010, Vol.1, pp. 537–541.
- [4] Yoon WJ., Cho YH., Park KS. A Study of Speech Emotion Recognition and Its Application to Mobile Services, ser. Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2007, vol 4611.
- [5] Gupta, P., Rajput, N. «Two-stream emotion recognition for call center monitoring», in *Proc. Interspeech 2007*, pp.2241–2244.
- [6] Szwoch, M., Szwoch, W. «Emotion recognition for affect aware video games», in *Image Processing & Communications Challenges 6*, Springer International Publishing, Cham, vol. 313, pp. 227–236.
- [7] Lancker, D.V., Cornelius, C., Kreiman, J. «Recognition of emotionalprosodic meanings in speech by autistic, schizophrenic, and normal children». *Develop. Neuropsychol.* vol. 5 (2–3), pp. 207–226, 1989.
- [8] Low, L.A., Maddage, N.C., Lech, M., Sheeber, L.B., Allen, N.B. (2011) «Detection of clinical depression in adolescents' speech during family interactions», *IEEE Trans. Biomed. Eng.* vol.58, issue 3, pp. 574–586.
- [9] Ververidis, D., Kotropoulos, C. «Emotional Speech Recognition: Resources, Features, and Methods», *Speech Communication*, vol.48, issue 9, pp. 1162–1181, 2006, [Online]. Available: <http://dx.doi.org/10.1016/j.specom.2006.04.003>
- [10] Ayadi, M.E., Kamel M.S., Karray F. «Survey on speech emotion recognition: Features, classification schemes, and databases», *Pattern Recognition*, vol. 44, issue 3, pp. 572–587, 2011.
- [11] Koolagudi, S. G., & Rao, K. S. «Emotion recognition from speech: a review», *International Journal of Speech Technology*, vol.15 issue 2, pp. 99–117, 2012.
- [12] Anagnostopoulos, C.N.; Iliou, T.; Giannoukos, I. «Features and classifiers for emotion recognition from speech: A survey from 2000 to 2011», *Artif. Intell. Rev.*, vol. 43, pp. 155–177, 2012.
- [13] Ramakrishnan, S. Recognition of emotion from speech: A review. In: Ramakrishnan, S. (Ed.), *Speech Enhancement, Modeling and Recognition Algorithms and Applications*, Intec, 2012.
- [14] Sailunaz, K., Dhaliwal, M., Rokne, J., Alhadj, R. «Emotion detection from text and speech: a survey» *Soc. Netw. Anal. Min.* 8(1), pp.1–26, 2018.
- [15] Basu, S., Chakraborty, J., Bag, A., Aftabuddin, M. «A review on emotion recognition using speech», in *International Conference on Inventive Communication and Computational Technologies (ICICCT)*, 2017, pp. 109–114.
- [16] Livingstone SR, Russo FA (2018) «The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English». *PLoS ONE* 13(5): e0196391. [Online]. Available: <https://doi.org/10.1371/journal.pone.0196391>. M. K. P. Kate Dupuis, «Toronto emotional speech set (TESS)» 2010. [Online]. Available: <https://tspace.library.utoronto.ca/handle/1807/24487>

Стаття надійшла до редколегії

26.07.2021



Intelligent module for recognizing emotions by voice

Speech is the main way of communication for people, and people can receive not only semantic but also emotional information from speech. Recognition of emotions by voice is relevant to areas such as psychological care, security systems development, lie detection, customer relationship analysis, video game development. Because the recognition of emotions by a person is subjective, and therefore inexact and time consuming, there is a need to create software that could solve this problem. The article considers the state of the problem of recognizing human emotions by voice. Modern publications, the approaches used in them, namely models of emotions, data sets, methods of extraction of signs, classifiers are analyzed. It is determined that existing developments have an average accuracy of about 0.75. The general structure of the system of recognition of human emotions by voice is analyzed, the corresponding intellectual module is designed and developed. A Unified Modeling Language (UML) is used to create a component diagram and a class diagram. RAVDESS and TESS datasets were selected as datasets to diversify the training sample. A discrete model of emotions (joy, sadness, anger, disgust, fear, surprise, calm, neutral emotion), MFCC (Mel Frequency Cepstral Coefficients) method for extracting signs, convolutional neural network for classification were used.. The neural network was developed using the TensorFlow and Keras machine learning libraries. The spectrogram and graphs of the audio signal, as well as graphs of accuracy and recognition errors are constructed. As a result of the software implementation of the intelligent module for recognizing emotions by voice, the accuracy of validation has been increased to 0.8.

Key words: recognition of emotions by voice, neural networks, deep learning, convolutional neural networks.



Олег Паріонов. Завідувач кафедри інтелектуальних технологій Київського національного університету імені Тараса Шевченка, канд. техн. наук, доцент, Київ, Україна. Напрями досліджень: засоби і технології для розроблення інформаційних систем управління технологічними процесами й об'єктами різної фізичної природи.

Oleg Parionov. Candidate of Technical Sciences, Associate Professor, Head of the Department of Intellectual Technologies, Faculty of Information Technologies, Taras Shevchenko National University of Kyiv, Kyiv, Ukraine. Areas of research - tools and technologies for developing information systems for managing technological processes and objects of different physical nature information systems.



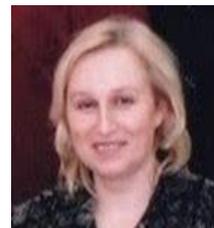
Ганна Красовська. Працює на посаді доцента кафедри інтелектуальних технологій Київського національного університету імені Тараса Шевченка, канд. техн. наук, Київ, Україна. Наукові інтереси: інтелектуальні системи підтримки прийняття рішень, адаптивні інтелектуальні системи в освіті, мультиагентні системи та технології.

Anna Krasovska. Candidate of Technical Sciences, Associate Professor, works as an associate professor of the Department of Intellectual Technologies of Taras Shevchenko National University of Kyiv, Kyiv, Ukraine. Research interests include intelligent decision support systems, adaptive intelligent systems in education, multi-agent systems and technologies.



Антон Астахов. Випускник магістерської освітньої програми «Технології штучного інтелекту» Київського національного університету імені Тараса Шевченка, Київ, Україна.

Anton Astakhov Graduate of the educational program "Technologies of Artificial Intelligence" of Taras Shevchenko National University of Kyiv, Kyiv, Ukraine.



Ірина Доманецька. Працює на посаді доцента кафедри інтелектуальних технологій Київського національного університету імені Тараса Шевченка, канд. техн. наук, Київ, Україна. Сфера наукових досліджень: системотехнічні дослідження в галузі ІТ, нейромережні технології та їхнє застосування, адаптивні системи навчання.

Iryna Domanetska. Candidate of Technical Sciences, Associate Professor, works as an associate professor of the Department of Intellectual Technologies of Taras Shevchenko National University of Kyiv, Kyiv, Ukraine. Areas of research are system-technical research in the field of IT, neural network technologies and their application, adaptive learning systems.



УДК 004.932

DOI: <https://doi.org/10.17721/AIT.2021.1.07>

I. А. Юрчук, orcid.org/0000-0001-8206-3395,
О. С. Колесник, orcid.org/0000-0001-8539-9310,
Київський національний університет імені Тараса Шевченка, Київ, Україна

Сегментація як ефективний метод виокремлення пухлини головного мозку на зображенні МРТ

Оброблення цифрових зображень, що виникають у багатьох сферах життя, є одним із напрямів, що вимагає стрімкого розвитку та вдосконалення існуючих алгоритмів як з метою точності та повноти, так і з міркувань швидкодії та економічності технічних і програмних рішень. Саме медичне застосування є тією галуззю, де важлива і точність оброблення, оскільки нестача інформації впливає на протокол лікування, і вартість, яка актуальна для доступності та широкого застосування. У межах цієї роботи запропоновано алгоритм сегментації цифрових зображень МРТ головного мозку з метою виокремлення сегмента, який містить пухлину. Указаний алгоритм базується на послідовному виконанні таких кроків: порогового методу Отцу бінаризації зображення, виокремлення тканин мозку та пухлини за допомогою морфологічних операцій, сегментації методом маркованого вододілу, видалення лінії черепа і виділення сегмента, що містить пухлину за допомогою ерозії.

У результаті верифікації не виявлено хибно-позитивних результатів сегментації, а відсоток зображень, що правильно відсегментовані з метою виявлення пухлини, складає 96,2 %. Необхідно відмітити високу швидкодію процесу сегментації, що отримана авторами.

Ключові слова – МРТ, сегментація, морфологічні методи, метод Отцу.

Для цитування (for citation): I. А. Юрчук, О. С. Колесник. “Сегментація як ефективний метод виокремлення пухлини головного мозку на зображенні МРТ” *Сучасні інформаційні технології*, vol. 1, p. 53–58, 2021.

ВСТУП

У розпізнаванні тих локацій головного мозку, де виникають аномалії на цифрових зображеннях, зокрема МРТ, комп'ютерній системі належить значна роль у полегшенні роботи медичних закладів, адже вона спрощує процес виявлення непомітних ореолів пухлин на початковому етапі хвороби, коли людське око майже не здатне відрізнити невеликі відмінності у кольорах. Аналізуючи велику кількість даних (цифрових зображень), комп'ютер може з точністю до 99,9 % визначити наявність пухлин головного мозку на цифрових зображеннях МРТ.

Оскільки оброблення даних магнітного резонансу спеціалістами є доволі трудомістким завданням, то розроблення методів автоматичної сегментації пухлин головного мозку залишається одним із пріоритетних і найскладніших завдань оброблення медичних даних. Точна сегментація може покращити діагностику, наприклад, оцінку розміру пухлини в часі. Важливою складовою такого процесу є отримання звіту. Адже він може

використовуватися з метою візуального контролю якості проведеної сегментації, а також можливості індивідуального динамічного порівняння результатів.

На сучасному етапі реалізовано декілька пакетів, проте кожен із них створено з конкретною вузькою метою. Наприклад, сегментація анатомічних структур головного мозку відбувається за допомогою пакетів FreeSurfer та FSL, а пакет FreeSurfer створювався для сегментації сірої та білої речовин. Необхідно зазначити, що вартість обладнання, яке має додаткові функції, що зменшують людський фактор в аналізі даних, зростає в рази. Тому програмні рішення, які на основі даних, отриманих за допомогою базового технічного обладнання (недорогого МРТ-сканера), можуть забезпечити точну діагностику, зокрема і на ранніх стадіях захворювання, зменшивши затрати людських ресурсів (медичних працівників), є затребуваними.

© Юрчук І. А., Колесник О. С., 2021



Мета дослідження – розробити алгоритм сегментації зображень МРТ головного мозку, який дозволить виявити наявність сегмента, що є пухлиною (містить аномальні клітини). Цей метод має бути точним і швидкодіючим, у програмній реалізації, що дасть можливість його використання як передоброблення для складніших систем (проведення хірургічних втручань тощо).

Завдання – провести огляд існуючих методів сегментації, розробити алгоритм, який забезпечує точну та швидку сегментацію цифрового зображення МРТ головного мозку з метою виокремлення пухлини, здійснити його програмну реалізацію та порівняти з існуючими рішеннями.

ОГЛЯД МЕТОДІВ СЕГМЕНТАЦІЇ

Серед ефективних методів сегментації виокремимо такі: сегментація на основі згорткових нейронних мереж, метод К-середніх, порогові методи, морфологічні методи, сегментація вододілами та сегментація на основі дискретних структур.

У роботах [1–3] представлено реалізацію сегментації МРТ головного мозку на основі згорткових нейронних мереж. Це підхід вимагає великих датасетів зображень (кількість зображень, що необхідні для навчання, має бути не меншою ніж кількість вільних параметрів), тривалої підготовки зображень і великих обчислювальних ресурсів, проте точність, з якою виокремлюється сегмент, що містить пухлину, є високою.

Менші затрати на обчислення мають процеси сегментації, що базуються на К-середніх, проте вони вимагають препроцесингу чи постпроцесингу зображень. У роботі [4] автори, усуваючи недолік попереднього задання кількості сегментів, проводять передоброблення зображення з метою локалізації пухлин, після чого кількість сегментів для роботи алгоритму К-середніх рівна двом у кожній із виокремлених локацій. Інший підхід запропоновано авторами у [5], де результат сегментації реалізований К-середніми, опрацьовують морфологічними операторами для отримання точніших результатів.

Порогові методи вимагають найменшого часу для оброблення і виділення сегментів, проте основним недоліком є необхідність у додатковій інформації, для прикладу у роботі [6] використовується оброблення фільтрами й видалення зображення черепа з початкового зображення [7] з метою коректного визначення значень.

Високу точність сегментування дають методи, що базуються на текстурних підходах [8], чи побудові складних дискретних структур [9], проте вони вимагають великих обчислювальних

потужностей, що у свою чергу впливає на вартість і доступність продуктів, створених на їхній основі.

ПОСТАНОВКА ЗАДАЧІ

Нехай задано цифрове зображення $I = \{p_{ij}\}$, де p_{ij} – піксель, який характеризується координатами (i, j) . Утворимо скінченний набір множин $\{S_k\}_{k=1}^N$ таких, що $\bigcup_{k=1}^N S_k = I$ і $S_{k_1} \cap S_{k_2} = \emptyset$ при $k_1 \neq k_2$. Знайдемо підмножину індексів $t = \{t_1, \dots, t_l\}$ таку, що довільні два пікселі p_{ij} та $p_{i'j'}$, де p_{ij} належить деякій множині S_q , $q \in t$, а $p_{i'j'}$ належить S_m , $m \notin t$, мають різні значення кольорових складових.

Необхідно зауважити, що у випадку, коли довільні два пікселі p_{ij} та $p_{i'j'}$, які p_{ij} належать деяким множинам S_q та S_l , де $q, l \in t$, то вони мають однакові значення кольорових складових. Цей висновок зроблено на основі того, що клітини головного мозку, які є аномальними (пухлиною), мають той самий електромагнітний відгук під час проведення МРТ.

Вхідними даними є підготовлений датасет (локації з аномаліями виділено спеціалістами у діагностиці пухлин головного мозку, що виявлені на МРТ), який складається зі 155 цифрових зображень. Необхідно автоматизовано сегментувати наявні зображення так, щоб існував сегмент, який складається лише з пікселів, які відповідають клітинам пухлини, за умов малих затрат у часі та за відсутності великих обчислювальних потужностей.

АЛГОРИТМ НА ОСНОВІ МОРФОЛОГІЧНИХ ОПЕРАЦІЙ, ВОДОДІЛУ ТА ПОРОГІВ

Морфологічні методи досліджують зображення за допомогою невеликої форми або шаблону. Цей елемент структурування розташовують у всіх пікселях зображення і порівнюють із відповідним околком пікселів. Деякі операції перевіряють, чи «вписується» елемент в окіл, у той час як інші перевіряють, «потрапляє» він в окіл або перетинає його за допомогою таких логічних операцій: NOT (логічне «НЕ»); AND (логічне «І»); OR (логічне «АБО»); XOR (виключає «АБО»)[10–11].

Прості методи порогового оброблення включають заміну кожного пікселя в зображенні чорним пікселем, якщо інтенсивність зображення $src(x, y)$ менша деякої фіксованої константи T (тобто $src(x, y) < T$), або білого пікселя, якщо інтенсивність зображення більша цієї константи. Основні з них: граничний метод із глобальним порогом і граничний метод з адаптивним порогом.



Під час сегментації вододілами зображення розглядають як топографічний ландшафт із гребенями і долинами. Висоти ландшафту зазвичай визначають значеннями сірих відповідних пікселів або величиною їхнього градієнта. Грунтуючись на такому тривимірному представленні, перетворення вододілу розбиває зображення на водозбірні басейни. Для кожного локального мінімуму водозбірний басейн містить усі точки, шлях найкрутішого спуску яких закінчується в мінімумі. Вододіли відокремлюють басейни один від одного. За наявності зашумлених даних медичних зображень виникає велика кількість невеликих ділянок. Ця проблема відома як «надмірна сегментація». Широко поширеними є вододіл шляхом дилатації та вододіл із маркерами.

Для запобігання надмірній сегментації використовують сегментацію вододілу, контрольовану маркерами. Попереднє оброблення зображення МРТ є основним етапом, який усуває шум і згладжує зображення. Щоб запобігти неправильній класифікації тканин мозку та немозкових тканин, проводиться видалення із зображення кісток черепа. Сегментація зображень здійснюється за допомогою маркованої сегментації вододілу. Потім із сегментованого зображення за допомогою морфологічної операції визначають область пухлини. Після цього знаходять місце розташування пухлинної області. Загальну схему алгоритму зображено на рис. 1.

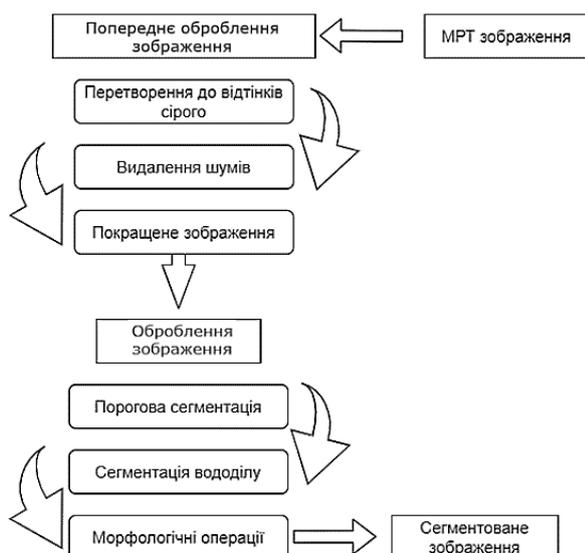


Рис. 1. Загальна схема реалізації процесу сегментації

Оскільки череп на МРТ-зображеннях виглядає як одна з найяскравіших областей зображення і різко контрастує з іншими областями мозку, то для того, щоб обмежити

виникнення хибнопозитивних сегментів, череп має бути вилючений із зображення.

Техніки видалення черепа можна розділити на три основні категорії [11]:

- **методи, засновані на інтенсивності**, базуються на пороговій класифікації. Основним недоліком цього підходу є його значна чутливість до коливань інтенсивності інтенсивності (у разі, наприклад, неоднорідності магнітного поля, що викликано зареєстрованим шумом або навіть властивостями пристрою);

- **методи, засновані на морфології**, мають основну ідею в тому, щоб об'єднати використання морфологічних операцій, порогових значень і методів виявлення країв, щоб найточніше відокремити область мозку від навколишньої тканини.

- **методи на основі деформованої моделі**, в якій застосовується деформація активного контуру й підгонка для локалізації областей мозку та її ідентифікація за допомогою характеристик зображення.

Алгоритм, представлений у цій роботі, належить до групи методів, заснованих на морфології:

- порогове оброблення зображень;
- заповнення прогалін у витягнутих об'єктах за допомогою морфологічних операторів;
- виявлення країв і поліпшення країв, якщо це необхідно;
- виокремлення найбільшої області зображення і створення бінарної маски;
- об'єднання двійкової маски і вхідного зображення як вихідного зображення.

Метод Отцу шукає поріг, що зменшує дисперсію всередині класу, яка визначається як зважена сума дисперсій двох класів:

$$\sigma_{\omega}^2(t) = \omega_1(t)\sigma_1^2(t) + \omega_2(t)\sigma_2^2(t),$$

де ваги ω_i – це ймовірності двох класів, розділених порогом t , σ_i^2 – дисперсія цих класів.

Алгоритм методу Отцу:

Нехай дано монохромне зображення $G(i, j), i = 1, \text{Height}, j = 1, \text{Width}$. Лічильник повторів $k=0$.

1. Побудувати гістограму $p(l)$ зображення та частоту $N(l)$ для кожного рівня інтенсивності зображення G .

2. Обчислити початкові значення для $\omega_1(0), \omega_2(0)$ та $\mu_1(0), \mu_2(0)$.

3. Для кожного значення $t = \overline{1, \max(G)}$ – полутону – горизонтальна вісь гістограми:

1. Оновлюємо ω_1, ω_2 та μ_1, μ_2 .

2. Обчислюємо

$$\sigma_b^2(t) = \omega_1(t)\omega_2(t)[\mu_1(t) - \mu_2(t)]^2.$$

3. Якщо $\sigma_b^2(t)$ більше, ніж наявне, то запам'ятовуємо σ_b^2 і значення порога t .

4. Шуканий поріг відповідає максимуму $\sigma_b^2(t)$:

$$N_T = \sum_{i=0}^{\max(G)} p(i),$$
$$\omega_1(t) = \frac{\sum_{i=0}^{t-1} p(i)}{N_T} = \sum_{i=0}^{t-1} N(i), \quad \omega_2(t) = 1 - \omega_1(t),$$
$$\mu_T = \frac{\sum_{i=0}^{\max(G)} i * p(i)}{N_T} = \sum_{i=0}^{\max(G)} i * \omega_1(t),$$
$$N(i), \mu_1(t) = \frac{\sum_{i=0}^{t-1} i * p(i)}{N_T * \omega_1(t)} = \frac{\sum_{i=0}^{t-1} i * N(i)}{\omega_1(t)},$$
$$\mu_2(t) = \frac{\mu_T - \mu_1(t) * \omega_1(t)}{\omega_2(t)}.$$

Алгоритм виокремлення сегмента з пухлиною:

1. Виконати бінаризацію Отцу;
2. Видалити невеликі білі шуми на зображенні за допомогою морфологічних операцій відкриття та закриття. Тоді область поруч із центром об'єкта є переднім планом, а область далеко від об'єкта – фоном.

3. За допомогою морфологічної операції ерозії видаляємо межові пікселі.

ОСОБЛИВОСТІ РЕАЛІЗАЦІЇ ПРОГРАМНОГО ЗАСТОСУНКУ

Програмний застосунок, що реалізує сегментацію цифрових зображень МРТ головного мозку, розроблено в інтегрованому середовищі розроблення PyCharm мови програмування python з використанням фреймворку Qt. Мінімальні системні вимоги: 64-бітна операційна система Windows, Python 3.5 або вище та 36 кБ доступного простору на жорсткому диску.

Вхідні дані:

- цифрові зображення форматів JPEG (.jpg) та PNG (.png);
- параметри: шляхи збереження файла/файлів.

Вихідні дані: сегментовані цифрові зображення МРТ головного мозку.

Далі визначаємо поріг і перетворюємо зображення на бінарне за допомогою порогового методу Отцу. Після порогового оброблення зображення, наступним кроком є створення маски, яка, помножена на вхідне зображення, дозволяє виділити тільки область тканини мозку. Використовуючи фільтр Кенні для виявлення меж, програма за допомогою методу зв'язних компонент виділяє найбільшу область зображення і робить маску, яку поєднує з початковим зображенням для отримання мозку без черепної коробки.

Зображення тканини мозку проходить фільтрацію шуму Гаусса, і за допомогою морфологічних операцій розширення та відкриття

виділяє області фону (тканини мозку) та об'єкта (пухлини) для проведення сегментації методом маркованого вододілу. Области, що не належать ні до множини фону, ні до множини об'єкта, визначаються, як маркери. Вони використовуються для проведення остаточної морфологічної операції – ерозії, що дозволяє виділити пухлину як окремий об'єкт.

Результатом роботи є виведений на екран масив зображень, що відображають поетапне оброблення зображення, з фінальним результатом – сегментованим зображенням, де область пухлини позначена різними відтінками сірого кольору. Програма дозволяє обрати користувачу велику кількість зображень для проведення сегментації, а також записати результати сегментації у документ Microsoft Word.

Приклад результату оброблення наведено на рис. 2. Зображення, які розташовані ліворуч, є результатами виконання основних етапів алгоритму, які за потреби нададуть фахівцеві з діагностики додаткову інформацію. Наприклад, зображення з видаленим черепом дає можливість візуально оцінити відношення площі сегмента, який містить пухлину, до площі головного мозку (череп видалено). На зображенні, що представлено праворуч на рис. 2, виокремлено декілька сегментів: сегмент, який складається з клітин пухлини, сегмент, який є «межовим» із пухлиною, та тим, що містить «здорові клітини» головного мозку, та інші.

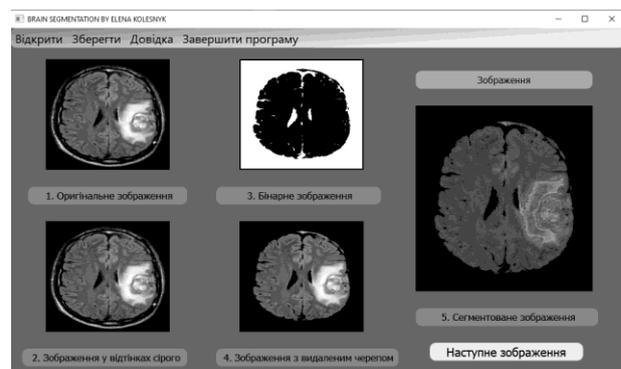


Рис. 2. Результат оброблення МРТ: різні сегменти представлено різними відтінками сірого

ЕКСПЕРИМЕНТАЛЬНА ВЕРИФІКАЦІЯ

Верифікація процесу сегментації була реалізована на 155 цифрових зображеннях МРТ головного мозку, що опрацьовані спеціалістами з метою отримання локалізації пухлин у структурах головного мозку. У середньому тривалість оброблення одного зображення склала 22 мілісекунди процесором Intel(R) Core



(TM) 5i – 4310U CPU 2.00GHz 2.60GHz та оперативною пам'яттю 4 ГБ. Загалом процес оброблення 155 зображень склав 43 секунди. Варто зауважити, що для порівняння, процес сегментації текстурним методом складає близько 30 секунд для одного зображення такої ж розмірності.

Для порівняння, авторами у роботі [12] отримано час оброблення 8 секунд нечітким С-середнім та 22 секунди К-середнім.

У результаті сегментації наявних зображень із хибно-позитивними результатами не було. Серед 155 зображень, на 6 зображеннях не було локалізовано пухлину (надано сегмент), що складає 3,8 % загальної кількості.

Також у межах дослідження проведено порівняння цього алгоритму з бібліотечними рішеннями OpenCV. Деякі результати порівняння сегментації зображено на рис. 3.

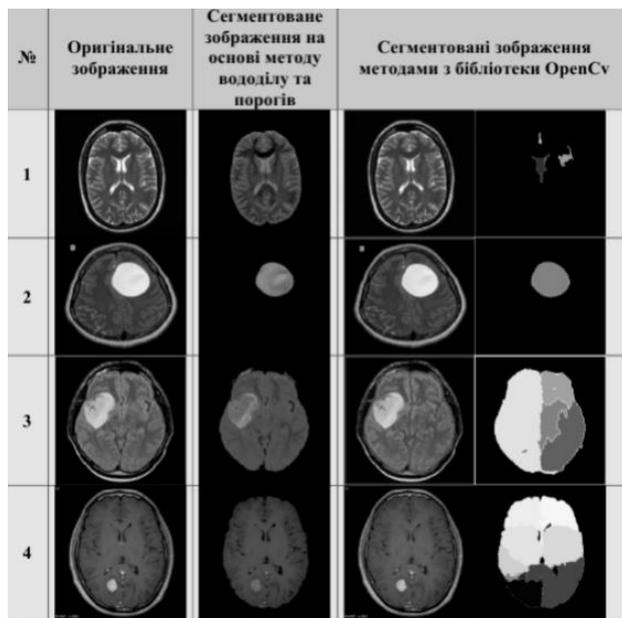


Рис. 3. Результати порівняння з бібліотечними рішеннями

У випадку 1 є хибне спрацювання, оскільки разом із сегментом, що містить пухлину, було виділено сегменти із здоровими клітинами. У зображених на рисунку випадках 3 та 4 готове бібліотечне рішення не виділило сегменти з пухлиною. Аналізуючи отримані дані, можна зробити такий висновок: не дивлячись на широкі можливості бібліотеки OpenCV і легкість її використання, сегментація деяких типів зображень видає некоректні результати. У випадку методу, який реалізовано в межах дослідження, комбінація методів порогової сегментації та вододілу в певній послідовності та

повторюваності (з різними вхідними масивами) дає кращий результат. Методи бібліотеки OpenCV хоч і створені для оброблення зображень, але не повністю пристосовані до сегментації саме МРТ-зображень мозку.

ВИСНОВКИ

Авторами досліджено методи сегментації зображення й отримано алгоритм сегментації цифрових зображень МРТ головного мозку з метою виділення сегмента, що містить пухлину. Запропонований алгоритм на основі почергового виконання бінаризації Отцу, сегментації маркованим вододілом та видалення черепа за допомогою морфологічних операцій, показав кращі результати порівняно з бібліотечними рішеннями та високу швидкодію.

У подальших дослідженнях необхідно вдосконалити механізм виявлення порога (зробити його чутливішим) з метою збільшення відсотка позитивно сегментованих зображень.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

- [1] R. Vinoth, and C. Venkatesh, "Segmentation and Detection of Tumor in MRI images Using CNN and SVM Classification," in Proc. ICEDSS, 2018, pp. 21–25.
- [2] M. Rahimpour, K. Goffin, and M. Koole, "Convolutional Neural Networks for Brain Tumor Segmentation Using Different Sets of MRI Sequences," in Proc. Nuclear Science Symposium and Medical Imaging Conference, 2019, pp. 1–3.
- [3] S. Somasundaram, and R. Gobinath, "Current Trends on Deep Learning Models for Brain Tumor Segmentation and Detection – A Review," in Proc. COMITCon, 2019, pp. 217–221.
- [4] T. Imtiaz, S. Rifat, and S. A. Fattah, "Automated Brain Tumor Segmentation from MRI Data Based on Local Region Analysis," in Proc. BECITHCON, 2019, pp. 107–110.
- [5] M. Kumar, A. Sinha, and N. V. Bansode, "Detection of Brain Tumor in MRI Images by Applying Segmentation and Area Calculation Method using SCILAB," in Proc. ICCUBEA, 2019, pp. 1–5.
- [6] A. Hussain, and A. Khunteta, "Semantic Segmentation of Brain Tumor from MRI Images and SVM Classification using GLCM Features," in Proc. ICIRCA, 2020, pp. 38–43.
- [7] T. M. S. Sazzad, K. M. T. Ahmed, M. Hoque, and M. Rahman, "Development of Automated Brain Tumor Identification Using MRI Images," in Proc. ECCE, 2019, pp. 1–4.
- [8] T. A. Jemimma, and Y. J. Vetharaj, "Watershed Algorithm based DAPP features for Brain Tumor Segmentation and Classification," in Proc. ICSSIT, 2019, pp. 155–158.
- [9] Yurchuk I., "Digital image segmentation based on the persistent homologies," in Proc. ICTES, 2019, pp. 226–232.
- [10] R.C. Gonzalez, Digital Image Processing, 2nd ed.; Prentice Hall: Upper Saddle River, NJ, USA, 2002.
- [11] R. Roslan, N. Jamil and R. Mahmud, "Skull Stripping of MRI Brain Images using Mathematical Morphology," in IEEE-EBMS Conference on Biomedical Engineering and Sciences (IECBES 2010), 2010, pp. 26–31.
- [12] B. Srinivas, and G. S. Rao, "Unsupervised learning algorithms for MRI brain tumor segmentation," in Proc. SPACES, 2018, pp. 181–184.

Стаття надійшла до редколегії

07.10.2021



Segmentation as an effective method of isolating a brain tumor on MRI

Digital image processing, which ensues in many sides of life, is one of the areas that requires rapid development and improvement of existing algorithms, both for accuracy and completeness, and for reasons of speed and cost-effectiveness of both technical and software solutions. Medical application itself is the area where both precision in processing is important, as insufficient information affects the treatment protocol, and the cost for availability and widespread use. In this research, an algorithm for segmentation of digital MRI images of the brain is proposed in order to isolate the segment that contains the tumor. This algorithm is based on the sequential execution of the following steps: threshold Otsu's method of binarization of the image, selection of brain and tumor tissues by morphological operations, segmentation by marked watershed, removal of the skull line and selection of the segment containing the tumor by an erosion.

The verification did not reveal false-positive segmentation results, and the percentage of images correctly segmented to detect the tumor was 96.2%. It should be noted the high speed of the segmentation process obtained by the authors.

Keywords: MRI, a segmentation, morphological methods, Otsu's method.



Ірина Юрчук. Працює доцентом кафедри програмних систем і технологій факультету інформаційних технологій, канд. фіз.-мат. наук, Київський національний університет імені Тараса Шевченка, Київ, Україна. Наукові інтереси: топологія многовидів малих розмірностей, топологічний аналіз даних та оброблення цифрових зображень.

Iryna Yurchuk. PhD in geometry and topology, associate professor of software systems and technologies, faculty of information technology. Taras Shevchenko National University of Kyiv, Kyiv, Ukraine. The author of publications on topology of low dimensional manifolds, topological data analysis and digital image processing.



Олена Колесник. Закінчила Київський національний університет імені Тараса Шевченка. Бакалавр з інженерії програмного забезпечення. Працює старшим лаборантом кафедри вищої математики Національного авіаційного університету, Київ, Україна. Займається розробленням алгоритмів сегментації цифрових зображень.

Olena Kolesnyk. Bachelor of software engineering. She graduated from Taras Shevchenko National University of Kyiv and works for National Aviation University as a senior laboratory assistant of higher mathematics department. Kyiv, Ukraine. She studies algorithms of digital images segmentation.



УДК 004.932.2

DOI: <https://doi.org/10.17721/AIT.2021.1.08>

М. В. Мороз, orcid.org/0000-0001-6953-683X,
Д. С. Берестов, orcid.org/0000-0002-3918-2978,
О. А. Курченко, orcid.org/0000-0002-3507-2392,

Київський національний університет імені Тараса Шевченка, Київ, Україна

Аналіз алгоритмів візуального відстежування об'єктів для систем реального часу

Проведено аналіз останніх досягнень і рішень у процесі візуального супроводу цільового об'єкта в галузі комп'ютерного зору, розглянуто підходи до вибору алгоритму візуального супроводження об'єктів на відеопослідовностях, виокремлено основні візуальні особливості, на яких може ґрунтуватись відстежування цільового об'єкта. Визначено критерії, які впливають на вибір алгоритму відстежування цільового об'єкта у режимі реального часу. Проте для відстежування у режимі реального часу за обмежених обчислювальних ресурсів питання вибору відповідного алгоритму є вирішальним. На вибір алгоритму візуального відстежування також впливають вимоги й обмеження щодо супроводжуваних об'єктів і попередні знання чи припущення щодо них. У результаті аналізу надано перевагу алгоритму відстежування Staple, за критерієм швидкодії, який є вирішальним показником у проектуванні та розробленні комплексу програмного та технічного забезпечення для автоматизованого візуального супроводу об'єкта у відеопотоці в режимі реального часу для різноманітних систем спостереження та безпеки, моніторингу трафіка дорожнього руху, розпізнавання діяльності й інших вбудованих систем.

Ключові слова: відеопоток, відстежувач, алгоритм, об'єкт, вбудована система, система реального часу.

Для цитування (for citation): М. В. Мороз, Д. С. Берестов, О. А. Курченко. "Аналіз алгоритмів візуального відстежування об'єктів для систем реального часу," *Сучасні інформаційні технології*, vol. 1, p. 59–65, 2021.

ВСТУП

Супроводження (відстеження, трекінг) об'єктів нині є досить важливим завданням у галузі комп'ютерного зору. Відстеження об'єктів має багато практичних застосувань [1–4], серед яких відеоспостереження, взаємодія між людиною та комп'ютером, навігація роботів, різноманітні системи спостереження та безпеки, моніторинг трафіка дорожнього руху, розпізнавання діяльності тощо.

Відстеження об'єктів націлено на те, щоб надати комп'ютеру можливість отримувати кращу модель реального світу. Оскільки комп'ютери за допомогою додаткових пристроїв не можуть бачити навколишнє середовище так само, як це робить людина, то відстеження об'єктів сприяє наближенню комп'ютерного сприйняття дійсності до людського, тим самим покращуючи взаємодію між людиною і комп'ютером.

ПОСТАНОВКА ЗАДАЧІ

Метою статті є огляд та аналіз наявних алгоритмів візуального супроводження об'єктів

на відеопослідовностях у режимі реального часу для вбудованих систем [5].

Проведений аналіз останніх досягнень і рішень [6–7] показав що у процесі візуального супроводу цільового об'єкта виокремлюють чотири важливі компоненти:

- 1) ініціалізацію цілі;
- 2) модель подання цілі на основі її особливостей;
- 3) прогноз руху;
- 4) пошук та отримання положення цілі.

Ініціалізація цілі полягає у виборі позиції об'єкта або певної зони навколо нього за допомогою таких допоміжних візуальних елементів: обмежувальна прямокутна зона, еліпс, центроїд, скелет фігури, контур, силует та ін.

Модель подання цілі зазвичай складається з певних візуальних особливостей, які притаманні об'єкту. Чим унікальнішими будуть особливості цілі, тим ефективнішим і точнішим буде виявлення. Вибір особливостей візуального представлення цільового об'єкта є важливим завданням, оскільки від них залежатиме ефективність і якість його супроводження.

© Мороз М. В., Берестов Д. С., Курченко О. А., 2021

Під час прогнозування руху в наступних кадрах оцінюється найбільш імовірне нове розташування цілі. На цьому етапі, зазвичай, використовують модель руху об'єкта. Це корисно у випадках, коли завчасно відомо про можливий рух цілі. Модель руху об'єкта може бути строго детермінованою, тобто, знаючи початкову точку руху, можна точно спрогнозувати розташування шуканого об'єкта в наступних кадрах. Також модель руху може мати локальну пам'ять, наприклад, якщо в декількох попередніх кадрах об'єкт рухався в одному напрямку, то зі значною вірогідністю, можна буде припустити, що в поточному кадрі тенденція руху в такому ж напрямку збережеться. Проте у випадках, коли рух цілі не детермінований або може стрімко змінюватись, використання моделі руху може бути надлишковим або повністю зайвим.

Під час пошуку цілі відбувається локалізація об'єкта на кадрі. За наявності прогнозування руху етап пошуку може залежати від його результату. У деяких випадках така взаємодія сприяє значному підвищенню продуктивності. Під час пошуку цілі, зазвичай, використовують різноманітні спеціалізовані детектори цілей.

Також наголошено на важливості гарної моделі подання цільового об'єкта і виокремлено такі основні візуальні особливості, на яких може ґрунтуватись відстежування об'єкта:

- а) колір;
- б) контури об'єктів;
- в) оптичний потік;
- г) текстура.

Зокрема представлено варіанти деяких підходів до розпізнавання, серед яких:

- а) точкові розпізнавачі;
- б) видалення фону;
- в) сегментація;

г) контрольоване навчання або навчання під наглядом;

- д) метод тимчасової різниці.

Також слід зазначити, що особливості супроводжувачів підтверджують комплексність та складність цієї царини. Можуть бути різні підходи та поєднання, які будуть надавати різні переваги і розв'язувати різні проблеми. Оскільки відстежування об'єктів є дослідницькою діяльністю, то кращим рішенням буде обрати певний алгоритм із кола доступних, адже розроблені алгоритми є надійними та чітко обґрунтованими.

ВИКЛАД ОСНОВНОГО МАТЕРІАЛУ

Алгоритми відстеження можна розділили на дві загальні категорії [8–9]: відстежувачі з використанням кореляційних фільтрів (ВКФ) та

відстежувачі без використання кореляційних фільтрів (ВБКФ).

Дискримінаційні фільтри кореляції активно використовують у різних галузях комп'ютерного зору, серед яких розпізнавання об'єктів, реєстрація зображень, перевірка обличчя та розпізнавання дій. У відстеженні об'єктів кореляційні фільтри (КФ) використовують для поліпшення надійності й ефективності.

Схеми відстеження на основі КФ виконують обчислення, послуговуючись частотними характеристиками для мінімізації обчислювальної вартості. Загальна архітектура цих алгоритмів застосовує підхід «відстеження за виявленням» і зображена на рис. 1. Фільтри кореляції ініціалізують спеціально вказаною й обмеженою зоною, у якій розташовується ціль, на початковому кадрі. Під час подальшого відстеження місце розташування цілі в новому кадрі оцінюється завдяки використанню попереднього положення цілі. Щоб ефективно представити зовнішній вигляд цілі, застосовують відповідні методи вилучення характеристик, які використовують для побудови мапи характеристик (мапи особливостей). Крайні межі зони згладжуються завдяки застосуванню косинусного фільтра. Згладжування відбувається через те, що основні характеристики цілі зосереджені ближче до центра, а межі можуть вносити додатковий шум або ж можуть містити об'єкти, які не належать цілі.

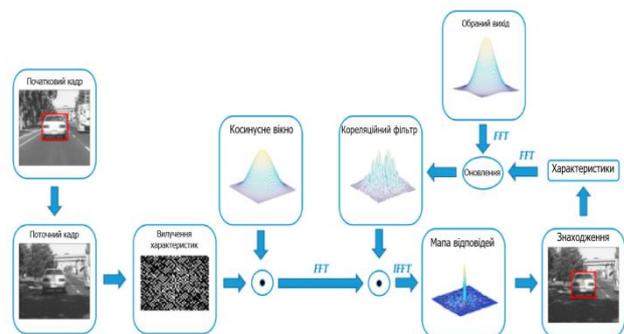


Рис. 1. Загальна схема алгоритмів відстеження на основі КФ

Операція кореляції виконується замість операції згортки, оскільки операція кореляції є швидшою. Мапа відповідей (результат пошуку в частотній області) обчислюється через поелементне множення між адаптивним фільтром навчання та характеристиками, вилученими за допомогою дискретної трансформції Фур'є (DFT). DFT працює із частотними даними за допомогою швидкого перетворення Фур'є (FFT). Мапу довіри (результат пошуку у просторовій області) отримують у просторовій області, застосовуючи



обернену FFT (IFFT) до мапи відповідей. Максимальне значення на мапі довіри вказує на нову позицію цілі. У результаті модель цілі оновлюється завдяки вилученню характеристик із нової позиції, після чого ці характеристики використовуються для оновлення кореляційного фільтра.

Схеми відстеження на основі КФ стикаються з різними труднощами, наприклад, із такими: можлива зміна із часом орієнтації та форми зовнішнього вигляду цілі. Іншим завданням є вибір ефективного представлення характеристик, оскільки деякі характеристики можуть покращувати ефективність ВКФ. Іншим важливим завданням для ВКФ є адаптація масштабу, оскільки розмір кореляційних фільтрів фіксований під час відстеження. Проте ціль із часом може змінювати свій розмір залежно від дальності до неї. Крім того, якщо ціль втрачено, її неможливо відновити знову.

Згідно із цією класифікацією всі відстежувачі, які не використовують кореляційні фільтри, називаються відстежувачами ВБКФ, які можна класифікувати на такі підкатегорії:

- відстежувачі з навчанням патчу;
- відстежувачі на основі розрідженості;
- відстежувачі на основі суперпикселів;
- відстежувачі на основі графів;
- відстежувачі на основі навчання з декількома екземплярами;
- відстежувачі на основі частин;
- відстежувачі на основі сіамських мереж.

Ці відстежувачі є дискримінаційними, за винятком відстежувачів на основі розрідженості, які є генеративними відстежувачами.

Відстежувачі з навчанням патчів використовують патчі цілі й патчі фону. Відстежувач навчається на позитивних і негативних зразках. Підготовлений відстежувач тестується на низці зразків, а максимальна відповідь відстежувача на одному з вхідних зразків вказує на положення цілі.

Під час *відстеження на основі навчання декількох екземплярів* замість розгляду певних патчів, як у попередньому методі, навчальні зразки розміщують у комірки й кожній комірці надається мітка. Позитивна мітка присвоюється комірці, якщо комірка містить принаймні один позитивний зразок, а негативна комірка містить лише негативні зразки. Позитивна комірка може містити позитивні та негативні екземпляри. Під час навчання мітки для екземплярів невідомі, але мітки на комірках відомі. За такого відстеження екземпляри використовують для побудови слабких класифікаторів, а кілька екземплярів вибирають та об'єднують для формування сильного класифікатора.

Відстежувач, що базується на сіамській мережі, здійснює відстеження на основі механізму відповідності. Процес навчання використовує загальні варіації зовнішнього вигляду цілі. Сіамські мережні відстежувачі узгоджують шаблони цілі зі зразками-кандидатами, щоб здобути схожість між патчами.

Відстежувачі на основі суперпикселів. Суперпикселі являють собою групу пикселів, що мають однакові значення. Цільова зона сегментується на декілька суперпикселів, що потім класифікуються для подальшої дискримінації об'єкта.

Відстежувачі на основі графів. Граф має вершини (це можуть бути пикселі, суперпикселі або частини об'єкта) та ребра (відповідність серед вершин). Графи використовують для прогнозування міток незазначених вершин. Як правило, відстежувачі на основі графів використовують суперпикселі як вузли для представлення зовнішнього вигляду об'єкта, а ребра представляють внутрішню геометричну структуру. Інша стратегія – побудувати графи між частинами об'єктів у різних кадрах.

Відстежувачі на основі частин. Частинне моделювання активно використовують для оброблення частин, які можуть змінюватись. Локальні частини об'єкта застосовують для моделювання трекера.

Відстежувачі на основі розрідженості. Усі вивчені до цього часу алгоритми є дискримінаційними методами відстеження. З іншого боку, генеративні методи вивчають представлення цілі та пошук цілі в кожному кадрі з мінімальною помилкою відновлення. Розріджене представлення – хороший приклад для генеративних моделей. Мета – виявити оптимальне подання цілі, яке є досить розрідженим і мінімізує помилку відновлення.

Для вибору алгоритму відстеження необхідно сфокусуватися на особливостях цільових об'єктів, особливостях середовища, обчислювальних можливостях пристрою та на перевагах наявних алгоритмів відстеження. Для більш зваженого рішення необхідно виконати порівняння наявних алгоритмів.

Щорічно спільнота «VOT Challenge» [10], діяльність якої спрямована на створення вільнодоступного сховища наявних алгоритмів візуального відстеження, проводить конкурс, у результаті якого порівнює і оцінює подані роботи. Це оцінювання служить для визначення якості відстежувача на основі підготовлених та анотованих відеопослідовностей. Тестові випадки включають основні проблеми відстеження, такі як: зміни освітлення, зміни розмірів



об'єкта, перекриття об'єкта тощо. Для автоматичного оцінювання якості відстежувачів використовують спеціально розроблений фреймворк. Також важливим є те, що у відкритому доступі перебувають як тестові дані з очікуваними результатами, так і результати порівняння трекерів за різними показниками.

Серед змагань «VOT Challenge» є також і змагання, які орієнтуються на відстежувачі, що придатні працювати в режимі реального часу. Здатність працювати в режимі реального часу означає, що відстежувач має повідомляти про розташування цілі на кожному кадрі із частотою, що перевищує або дорівнює частоті кадрів у відеопослідовності.

Спосіб тестування відстежувачів реального часу є таким. Спочатку відбувається ініціалізація відстежувача вхідними даними, а саме – цільовою зоною, у першому кадрі послідовності за допомогою відповідного інструментарію. Після цього інструмент оцінювання очікує вихідні дані відстежувача про положення цілі. Якщо відстежувач не встигне відповісти про нове положення цілі за наданий йому інтервал часу, то як відповідь відстежувача на такому кадрі беруть останню відповідь відстежувача.

Змагання «VOT Challenge» використовують методологію, що базується на перезапусках. У випадках, коли відстежувач надає результат у вигляді зони, в якій визначено нове розташування об'єкта, і ця зона має нульове перекриття з еталонними даними, то відбувається збій, а відстежувач повторно ініціалізується через п'ять кадрів після збою. Також у «VOT» є власні основні міри для оцінювання відстежувачів.

Точність і робастність – це основні міри, що застосовують для дослідження ефективності відстежувача в експериментах на основі перезапусків. Точність – це середнє перекриття між результатами відстежувача й еталонними даними впродовж успішних періодів супроводження. Робастність вимірює, як часто під час супроводу відстежувач втрачає ціль, унаслідок чого виконує повторну ініціалізацію. Потенційні відхилення в оцінюванні точності, які пов'язані з перезапусками, зменшуються завдяки ігноруванню десяти кадрів після повторної ініціалізації, що є досить консервативним запасом.

Третьою основною мірою оцінювання є очікуване середнє перекриття (ЕАО), яке поєднує значення точності в кожному кадрі та кількість перезавантажень відстежувача. Ця міра відповідає за очікуване середнє перекриття відстежувача на подібних за візуальними властивостями до такого набору даних короткотермінових послідовностях.

ЕАО вимірює очікуване перекриття зон відстежувача без перезавантажень на короткостроковій послідовності.

Для вибору й порівняння алгоритмів можна скористатися даними [11–14] щодо переможців у категорії реального часу за 2017–2020 рр. Порівняння алгоритмів наведено у таблиці, де є дані щодо оцінювання кожного алгоритму за значеннями середнього очікуваного перекриття, точності, робастності та використання алгоритмом обчислювальних ресурсів відеокарти (GPU) у процесі роботи. Також варто зазначити про трактування значень кожної з мір оцінювання. Кращими є більші значення очікуваного середнього перекриття, більші значення точності та менші значення робастності.

ТАБЛИЦЯ
Порівняння найкращих алгоритмів

Алгоритм	ЕАО	Точність	Робастність	GPU
AlphaRef	0,486	0,754	0,788	Так
OceanPlus	0,471	0,679	0,824	Так
AFOD	0,458	0,708	0,780	Так
SiamMargin	0,366	0,577	0,321	Так
SiamFCOT	0,350	0,601	0,386	Так
DiMP	0,321	0,582	0,371	Так
SiamRPN	0,383	0,586	0,276	Так
SA_Siam_R	0,337	0,566	0,258	Так
SA_Siam_P	0,286	0,533	0,337	Так
CSRDCF++	0,212	0,459	0,398	Так
SiamFC	0,182	0,502	0,604	Так
ECOhc	0,177	0,494	0,571	Ні
Staple	0,170	0,530	0,688	Ні

З наведеної таблиці видно, що збільшується тенденція до використання GPU та нейронних мереж у алгоритмах відстеження. Проте не всі вбудовані системи мають у своїй апаратній архітектурі GPU, а якщо й мають, то наявний GPU не завжди доцільний для виконання важких завдань. Наприклад, Raspberry Pi [15] хоч і має GPU, проте він не є достатньо потужним для виконання завдання відстежування і використовується, зазвичай, для завдань мультимедіа. Якщо ж реалізувати ці алгоритми без використання GPU, то їхня швидкість буде значно менша. Тому одним із критеріїв, який вплине на вибір алгоритму, є направленість алгоритму на виконання на CPU.

Для подальшого розгляду можна одразу відкинути алгоритми, які використовують GPU. Відповідно, залишилися лише алгоритми Staple та



ЕСОнс. Алгоритм ЕСОнс є кращим за показником робастності, а також має трохи кращий показник ЕАО. З іншого боку, Staple має кращий показник точності. Кожен із них запускався на процесорі Intel i7. ЕСОнс зумів досягти 60 кадрів у секунду, у той час як Staple – 80 кадрів у секунду, причому автори відстежувача Staple додатково вказали на модифікації, які збільшують швидкість відстежування до 100 кадрів у секунду. Саме тому перевага у виборі алгоритму відстежування буде надана алгоритму Staple, оскільки швидкодія є вирішальним показником у цьому випадку.

Алгоритм відстеження Staple у процесі роботи спирається на комбінацію фільтра кореляції (з використанням гистограми спрямованих градієнтів як характеристики) та глобальної кольорової гистограми. Він працює згідно з парадигмою відстежування на основі розпізнавання. Проблему пошуку об'єкта на зображенні x_t у порядку кадру t можна звести до задачі вибору такої прямокутної зони p_t з набору S_t , яка збільшуватиме оцінку у формулі

$$p_t = \arg \max_{p \in S_t} f(T(x_t, p), \theta), \quad (1)$$

де функція T – таке перетворення зображення, що $f(T(x, p), \theta)$ присвоює оцінку прямокутній зоні p на зображенні x відповідно до параметрів моделі θ ; θ – параметри моделі, які мають обиратися так, щоби мінімізувати значення функції витрат, яка залежить від попередніх зображень і положень об'єкта на цих зображеннях.

Для пошуку об'єкта на зображенні використовується оцінювальна функція. Оцінювальна функція є лінійною комбінацією шаблонної оцінки й оцінки гистограми і визначена як

$$f(x) = \gamma_{tmp} f_{tmp}(x) + \gamma_{hist} f_{hist}(x), \quad (2)$$

де γ_{tmp} , γ_{hist} – коефіцієнти, які визначають вплив шаблонної оцінки й оцінки гистограми на результуюче значення оцінки, причому

$$\gamma_{tmp} + \gamma_{hist} = 1.$$

Шаблонна оцінка є лінійною функцією характеристичного K -канального зображення φ_x , отриманого із зображення x та визначеного на скінченній сітці $\tau \subset Z^2$, обчислюється як

$$f_{tmp}(x; h) = \sum_{u \in \tau} h[u]^T \varphi_x[u], \quad (3)$$

де h – шаблон, інше характеристичне K -канальне зображення.

Оцінка гистограми є лінійною функцією середнього пікселя й обчислюється з характеристичного M -канального зображення ψ_x , отриманого із зображення x і визначеного на скінченній сітці $H \subset Z^2$:

$$f_{hist}(x; \beta) = \beta^T \left(\frac{1}{|H|} \sum_{u \in H} \psi_x[u] \right), \quad (4)$$

де β – ваговий вектор гистограми, шаблон.

Оскільки зовнішній вигляд об'єкта може суттєво змінюватися впродовж усієї відеопослідовності, то не ефективно будувати його модель лише на основі першого кадру та використовувати цю єдину модель, щоби знайти об'єкт у всіх інших кадрах. Тому цей алгоритм застосовує адаптацію моделі для того, аби скористатися інформацією, наявною в пізніших кадрах. Для цього він використовує прогнози (знайдене положення цілі) у нових кадрах як навчальні дані, завдяки яким можна оновити модель. Проте в цьому підході існує небезпека того, що невеликі помилки із часом накопичуються й у результаті можуть викликати дрейф моделі.

Отримати значення кореляційного фільтра (шаблону) для шаблонної оцінки можна за формулою

$$\hat{h}[u] = 1/(\hat{d}[u] + \lambda) \cdot \hat{r}[u], \quad (5)$$

де \wedge позначає операцію дискретного перетворення Фур'є; \hat{d} – знаменник шаблону, який визначено у формулі (6); \hat{r} – чисельник шаблону, який визначено у формулі (7); λ – коефіцієнт, що використовується для обмеження складності моделі.

$\hat{d}[u]$ є сумою елементів головної діагоналі матриці $\hat{s}[u]$ з розмірністю $K \times K$ з елементами $\hat{s}^{ij}[u] = (\hat{\phi}^j)^* \otimes \hat{\phi}^i$, де $\hat{\phi}$ є характеристичним K -канальним зображенням $\varphi_{T(x,p)}$, яке переведено до простору Фур'є. $\hat{d}[u]$ обчислюють за формулою

$$\hat{d}[u] = \sum_{i=1}^K (\hat{\phi}^i)^* \otimes \hat{\phi}^i, \quad (6)$$

де $*$ – комплексне спряжене число;

\otimes – операція поелементного множення.

$\hat{r}[u]$ є K -вимірним вектором з елементами $\hat{r}^i[u]$ й обчислюється за формулою

$$\hat{r}^i[u] = (\hat{y}^i)^* \otimes \hat{\phi}^i, \quad (7)$$

де y – бажаний вихід, зазвичай у формі функції Гаусса з піком в одиниці.

Формули (6) та (7) завдяки простору Фур'є представляють собою операції кроскореляції. Операція кореляції застосовується замість виснажливої щодо обчислювальних ресурсів операції згортки. Мапа відповідей обчислюється за допомогою поелементного множення між адаптивним кореляційним фільтром навчання й отриманими із зображення характеристиками за допомогою дискретної трансформації Фур'є



(DFT). DFT працює в частотній області за допомогою швидкого перетворення Фур'є (FFT). Мапа довіри у просторовій області отримується за допомогою оберненого FFT (IFFT) над мапою відповідей. Максимальні значення на мапі довіри вказують на нову цільову позицію об'єкта.

Навчання для шаблонної оцінки відбувається через адаптацію чисельника \hat{r}_t та знаменника \hat{d}_t до шаблону \hat{h} . Адаптація чисельника \hat{r}_t представлена формулою

$$\hat{r}_t = (1 - \eta_{mpl})\hat{r}_{t-1} + \eta_{mpl}\hat{r}_t', \quad (8)$$

де η_{mpl} – коефіцієнт адаптивності шаблонної моделі; \hat{r}_{t-1} – чисельник шаблону, отриманий із попередніх кадрів; \hat{r}_t' – чисельник шаблону, який обчислюється за формулою (7) для поточного кадру.

Адаптація знаменника \hat{d}_t представлена формулою

$$\hat{d}_t = (1 - \eta_{mpl})\hat{d}_{t-1} + \eta_{mpl}\hat{d}_t', \quad (9)$$

де \hat{d}_{t-1} – знаменник шаблону, отриманий із попередніх кадрів; \hat{d}_t' – знаменник шаблону, який обчислюється за формулою (6) для поточного кадру.

Отримати шаблон кольорової гістограми для характеристичного каналу зображення $j = 1 \dots M$ можна за формулою

$$\beta_j^j = \frac{p^j(O)}{p^j(O) + p^j(B) + \lambda}, \quad (10)$$

де $p^j(O)$ – частка пікселів у регіоні об'єкта, для якої характеристика j не дорівнює нулю; $p^j(B)$ – частка пікселів у регіоні фону об'єкта, для якої характеристика j не дорівнює нулю.

Навчання моделі на основі гістограми відбувається за допомогою адаптації моделей зони фону та зони об'єкта. Адаптація зони об'єкта представлена формулою

$$p_t(O) = (1 - \eta_{hist})p_{t-1}(O) + \eta_{hist}p_t'(O), \quad (11)$$

де η_{mpl} – коефіцієнт адаптивності гістограмної моделі; $p_t(O)$ – вектор з елементів $p_t^j(O)$ для $j = 1 \dots M$; $p_t'(O)$ – кольорова гістограма зони об'єкта, обчислена у кадрі t .

Адаптація зони фону представлена формулою

$$p_t(B) = (1 - \eta_{hist})p_{t-1}(B) + \eta_{hist}p_t'(B), \quad (12)$$

де $p_t(B)$ – вектор з елементів $p_t^j(B)$ для $j = 1 \dots M$; $p_t'(B)$ – кольорова гістограма зони фону біля об'єкта, обчислена у кадрі t .

ВИСНОВКИ

Візуальне відстежування об'єктів є актуальною проблемою в межах комп'ютерного зору. Завдяки її важливості нині існує багато підходів для розв'язання цієї проблеми. Проте для відстеження у режимі реального часу за обмежених апаратних ресурсів питання вибору відповідного алгоритму є вирішальним. На вибір відстежувача також впливають вимоги й обмеження щодо супроводжуваних об'єктів і попередні знання чи припущення щодо них. Для вибору відстежувача проведено огляд наявних алгоритмів, у результаті якого обрано алгоритм Staple та обґрунтовано доцільність його використання. Цей алгоритм має добру швидкодію, а також помірні вимоги щодо обчислювальних ресурсів, що є вирішальним для роботи на вбудованих системах. Подальші дослідження будуть направлені на проектування та розроблення комплексу програмного та технічного забезпечення для автоматизованого візуального супроводу об'єкта у відеопотоці в режимі реального часу для вбудованих систем.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

- [1] Y. Zhang, Z. Chen and B. Wei, "A Sport Athlete Object Tracking Based on Deep Sort and Yolo V4 in Case of Camera Movement", in *IEEE 6th International Conference on Computer and Communications (ICCC)*, 2020, pp. 1312–1316.
- [2] (2021) Оптико-електронна система керування стрільбою артустановок "Sens-2". [Online]. Available: <https://www.spetstechnoexport.com/product/sens-2>.
- [3] Byung-Kuk Seo, Jungsik Park and Jong-Il Park, "3-D visual tracking for mobile augmented reality applications", in *IEEE International Conference on Multimedia and Expo*, 2011, , pp. 1–4.
- [4] Mahendra Mallick; Vikram Krishnamurthy; Ba-Ngu Vo, "Tracking and Data Fusion for Ground Surveillance", in *Integrated Tracking, Classification, and Sensor Management: Theory and Applications*, IEEE, 2012, pp.203–254.
- [5] Peter Marwedel. (2021) Embedded System Design. [Online]. Available: <https://link.springer.com/content/pdf/10.1007%2F978-3-030-60910-8.pdf>.
- [6] Z. Soleimanitaleb, M. A. Keyvanrad and A. Jafari, "Object Tracking Methods: A Review", in 9th International Conference on Computer and Knowledge Engineering (ICCKE), 2019, pp. 282–288.
- [7] Sanna Ågren. (2021) Object tracking methods and their areas of application: A meta-analysis. A thorough review and summary of commonly used object tracking methods. [Online]. Available: <http://www8.cs.umu.se/education/examina/Rapporter/SannaAgrenFinal.pdf>.
- [8] Mustansar Fiaz, Arif Mahmood, Sajid Javed, and Soon Ki Jung. (2021) Handcrafted and Deep Trackers: Recent Visual Object Tracking Approaches and Trends. [Online]. Available: <https://arxiv.org/pdf/1812.07368.pdf>.
- [9] Yilmaz A., Javed O., and Shah M. (2021) Object tracking: A survey. [Online]. Available: <https://dl.acm.org/doi/pdf/10.1145/1177352.1177355>.
- [10] (2021) The VOT website. [Online]. Available: <https://www.votchallenge.net>



- [11] (2017) The VOT website. [Online]. Available: <https://www.votchallenge.net/vot2017/results.html>
- [12] (2018) The VOT website. [Online]. Available: <https://www.votchallenge.net/vot2018/results.html>
- [13] (2019) The VOT website. [Online]. Available: <https://www.votchallenge.net/vot2019/results.html>
- [14] (2020) The VOT website. [Online]. Available: <https://www.votchallenge.net/vot2020/results.html>

- [15] L. Rosyidi, A. Prasetyo and M. S. Romadhon, "Object Tracking with Raspberry Pi using Histogram of Oriented Gradients (HOG) and Support Vector Machine (SVM)", in *8th International Conference on Information and Communication Technology (ICoICT)*, 2020, pp. 1–6.

Стаття надійшла до редколегії

17.10.2021

Analysis of visual object tracking algorithms for real-time systems

The article analyzes the latest achievements and decisions in the process of visual support of the target object in the field of computer vision, considers approaches to the choice of algorithm for visual support of objects on video sequences, highlights the main visual features that can be based on tracking object. The criteria that influence the choice of the target object-tracking algorithm in real time are defined. However, for real-time tracking with limited computing resources, the choice of the appropriate algorithm is crucial. The choice of visual tracking algorithm is also influenced by the requirements and limitations for the monitored objects and prior knowledge or assumptions about them. As a result of the analysis, the Staple tracking algorithm was preferred, according to the criterion of speed, which is a crucial indicator in the design and development of software and hardware for automated visual support of the object in real-time video stream for various surveillance and security systems, monitoring traffic, activity recognition and other embedded systems.

Keywords: video stream, tracker, algorithm, object, embedded system, real-time system.



Микола Мороз. Студент четвертого курсу кафедри програмних систем і технологій факультету інформаційних технологій Київського національного університету імені Тараса Шевченка, Київ, Україна. Наукові інтереси: Internet of Things, Embedded Systems.

Mykola Moroz. 4th-year student of the Department of Software Systems and Technologies, Faculty of Information Technology, Taras Shevchenko National University of Kyiv, Kyiv, Ukraine. Research interests: Internet of Things, Embedded Systems.



Олег Курченко. Доцент кафедри програмних систем і технологій факультету інформаційних технологій Київського національного університету імені Тараса Шевченка, Київ, Україна. Наукові інтереси: інженерія програмного забезпечення, технології віртуалізації, штучний інтелект.

Oleg Kurchenko. Associate Professor of the Department of Software Systems and Technologies, Faculty of Information Technology, Taras Shevchenko National University of Kyiv, Kyiv, Ukraine. Research interests: Software Engineering, Virtualization Technologies, Artificial Intelligence.



Денис Берестов. Асистент кафедри програмних систем і технологій факультету інформаційних технологій Київського національного університету імені Тараса Шевченка, Київ, Україна. Наукові інтереси: методи та інструменти Data Science, Industrial Internet of Things, Embedded Systems.

Denys Berestov. Assistant of the Department of Software Systems and Technologies, Faculty of Information Technology, Taras Shevchenko National University of Kyiv, Kyiv, Ukraine. Research interests: methods and tools of Data Science, Industrial Internet of Things, Embedded Systems.



ІНФОРМАЦІЙНА АНАЛІТИКА Й АНАЛІТИКА ДАНИХ



УДК 004.67

DOI: <https://doi.org/10.17721/AIT.2021.1.09>

Liudmyla Zubyk, orcid.org/0000-0002-2087-5379,
Taras Shevchenko National University of Kyiv, Kyiv, Ukraine,
Yaroslav Zubyk, orcid.org/0000-0002-0802-3552,
National University of Water and Environmental Engineering, Rivne, Ukraine

Architecture of modern platforms for big data analytics

Big data is one of modern tools that have impacted the world industry a lot of. It also plays an important role in determining the ways in which businesses and organizations formulate their strategies and policies. However, very limited academic researches has been conducted into forecasting based on big data due to the difficulties in capturing, collecting, handling, and modeling of unstructured data, which is normally characterized by its confidential. We define big data in the context of ecosystem for future forecasting in business decision-making. It can be difficult for a single organization to possess all of the necessary capabilities to derive strategic business value from their findings. That's why different organizations will build, and operate their own analytics ecosystems or tap into existing ones. An analytics ecosystem comprising a symbiosis of data, applications, platforms, talent, partnerships, and third-party service providers lets organizations be more agile and adapt to changing demands. Organizations participating in analytics ecosystems can examine, learn from, and influence not only their own business processes, but those of their partners. Architectures of popular platforms for forecasting based on big data are presented in this issue.

Keywords: big data, unstructured data, platforms for data analytic, data ecosystems, big data environment.

Для цитування (for citation): L. Zubyk, Y. Zubyk. "Architecture of modern platforms for big data analytics," *Сучасні інформаційні технології*, vol. 1, p. 67–74, 2021.

INTRODUCTION

The world is changing, and the variations in living conditions that we have all seen lately play a key role in just such a resource as information. Modern information technologies are already forming new structures and creating enormous opportunities for development. Finally, only those who can adapt to new conditions faster will be able to win. The amount of data is growing exponentially, a clear example is the amount of information generated over the past 2 years by users of social networks. Currently, the daily data flow is more than 8 TB, and this number is growing daily. Millions of photos, videos, texts, hundreds of terabytes of information are uploaded to the public every minute. Information content in the form of documents, books, movies is constantly digitized. IoT devices and sensors regularly update the system with information from their owners, which allows them to build appropriate behavioral models and use them for future development. IoT and machine learning are causing the global Big Data market grow rapidly, and is forecasted to 103 billion U.S. dollars by 2027 [1].

FORMULATION OF THE PROBLEM

The big data resource is universal, dynamic and inexhaustible. Due to the availability of significant amounts of collected data, high quality digital information and its availability in developed countries among the world leaders in using of Big Data were primarily the United States, Great Britain, China, Switzerland, South Korea etc. [2].

Among the world-famous companies working with Big Data: iTechArt, ScienceSoft, Xplenty, IBM, HP Enterprise, Teradata, Oracle, SAP, EMC, Amazon, Microsoft Google, VMware, Splunk, Alteryx, Cogito, etc. [3].

This direction is quite new for Ukraine. Among the projects that have gone far beyond the country – Grammarly (a system for constructing spell-checking algorithms based on Big Data). Both government agencies (the Ministry of Digital Transformation) and business representatives are interested in big data analytics.

Big Data in the world is used in almost all areas: medicine, telecommunications, logistics, urban planning, retail, energy, agriculture and finance, space exploration. Customers of services for the



analysis of large arrays of information are both large holdings and representatives of medium and small businesses. Powerful streaming platforms (Netflix, Spotify, etc.) analyze video and audio content; Amazon studies shopping history; Tinder creates individual portraits of users. Most Big Data projects focus on either expanding the customer base or developing existing customers and improving existing services [10; 19].

Large amounts of customer data can be available for analysis only with the direct participation of their holders, ie banks, retailers, mobile operators and so on. Data sets of Internet service users always provide a good basis for building analytical models.

The processing of personal data of customers places strict requirements on companies to comply with current legislation in terms of protection of personal data of users, so all analytical models using Big Data are based solely on impersonal data [4; 5].

The first Big Data cases for customers appeared a long time ago and since then many companies have entered the market, specializing in the development of specific platforms and solutions for customers. Big Data analysts and specialists work with an internal platform for data processing and analysis based on Open Source technologies, constantly creating new products and improving existing ones.

2020 and 2021 tested many companies for strength in the face of constant uncertainty, abrupt changes in the business environment and unusually severe constraints. Successful examples of those companies that have managed not only to stay afloat, but also to improve their business performance, include the network of fuel filling stations UPG, Brain, online store ITbox.ua and more.

Data analysis for business customers is often aimed at developing and involving in business development targeting programs to organize the influx of customers through bonus offer systems; construction of portraits of clients to determine the target groups and their needs; expanding markets by forming Look-Alike audiences; active use of heatmap and geoanalytics in order to optimize the integration of new objects into the infrastructure; organization of communication with existing and potential customers depending on the events that occur with them (trigger mailing); replacing traditional SMS with Viber messages, given that they may contain useful multimedia interactive content, etc. [20].

Large amounts of information are only part of success. The data flow detached from the analytics does not provide any advantages. For further use, the information needs to be cleaned and reworked. The growing popularity of Big Data is largely due to

changes in the technologies and infrastructure used for processing [17].

A data ecosystem is a collection of infrastructure, analytics, and applications used to capture and analyze data. Data ecosystems provide companies with data that they rely on to understand their customers and to make better decisions. Like real ecosystems, data ecosystems are intended to evolve over time. There is no one 'data ecosystem' solution. Every business creates its own ecosystem, sometimes referred to as a technology stack, and fills it with a patchwork of hardware and software to collect, store, analyse, and act upon the data.

METHODS OF THE RESEARCH

Methods of the research - empirical, or rather observation and comparison.

COMPARING OF DIFFERENT PLATFORMS ARCHITECTURE FOR DATA ANALYTICS

The best data ecosystems are built around a product analytics platform that ties the ecosystem together. Analytics platforms help teams integrate multiple data sources, provide machine learning tools to automate the process of conducting analysis, and track user cohorts so teams can calculate performance metrics.

The Hadoop Distributed File System (HDFS) is a distributed file system designed to run on commodity hardware. This software is suitable for applications that have large data sets. HDFS is highly fault-tolerant and is designed to be deployed on low-cost hardware. It provides high throughput access to application data. HDFS relaxes a few POSIX requirements to enable streaming access to file system data. A core architectural goal of HDFS is detection of faults and quick, automatic recovery from them. Software is designed more for batch processing. The emphasis is on high throughput of data access HDFS is tuned to support large files. HDFS applications need a write-once-read-many access model for files. A file once created, written, and closed need not be changed. This assumption simplifies data coherency issues and enables high throughput data access. A computation requested by an application is much more efficient if it is executed near the data it operates on. HDFS has been designed to be easily portable from one platform to another. It has a master/slave architecture (Fig. 1).

The NameNode and DataNode are pieces of software designed to run on commodity machines. The system is designed in such a way that user data never flows through the NameNode. HDFS supports a traditional hierarchical file organization.

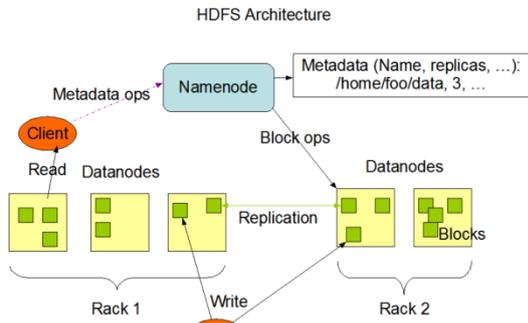


Fig. 1. HDFS Architecture

It doesn't yet implement user quotas. HDFS doesn't support hard links or soft links. However, the HDFS architecture doesn't preclude implementing these features.

An application can specify the number of replicas of a file that should be maintained by HDFS.

The number of copies of a file forms the replication factor of that file. This information is stored by the NameNode. It stores each file as a sequence of blocks; all blocks in a file except the last block are the same size. The blocks of a file are replicated for fault tolerance. The block size and replication factor are configurable per file. Optimizing replica placement distinguishes HDFS from most other distributed file systems. The NameNode uses a transaction log called the EditLog to persistently record every change that occurs to file system metadata.

All HDFS communication protocols are layered on top of the TCP/IP protocol.

The primary objective of HDFS is to store data reliably even in the presence of failures. The three common types of failures are NameNode failures, DataNode failures and network partitions.

DataNode sends a Heartbeat message to the NameNode periodically. The HDFS architecture is compatible with data rebalancing schemes. When a client creates an HDFS file, it computes a checksum of each block of the file and stores these checksums in a separate hidden file in the same HDFS namespace. The client can opt to retrieve that block from another DataNode.

NameNode can be configured to support maintaining multiple copies of the FsImage and EditLog. This may degrade the rate of namespace transactions per second that a NameNode can support. The NameNode machine is a single point of failure for an HDFS cluster. Currently, automatic restart and failover of the NameNode software to another machine is not supported. A typical block size used by HDFS is 64 MB. A client request to create a file does not reach the NameNode

immediately. In fact, initially the HDFS client caches the file data into a temporary local file. When the local file accumulates data worth over one HDFS block size, the client contacts the NameNode. The NameNode inserts the file name into the file system hierarchy and allocates a data block for it.

HDFS can be accessed from applications in many different ways. Natively, HDFS provides a Java API for applications to use. A C language wrapper for this Java API is also available. In addition, an HTTP browser can also be used to browse the files of an HDFS instance. Work is in progress to expose HDFS through the WebDAV protocol.

It provides a command-line interface called FS shell that lets a user interact with the data in HDFS. The DFSAdmin command set is used for administering an HDFS cluster.

When a file is deleted by a user or an application, it isn't immediately removed from HDFS.

When the replication factor of a file is reduced, the NameNode selects excess replicas that can be deleted.

Azure Databricks is a platform optimized for the Microsoft Azure cloud services platform. This platform includes such environments for developing applications: SQL Analytics and Workspace (Fig. 2) [13].

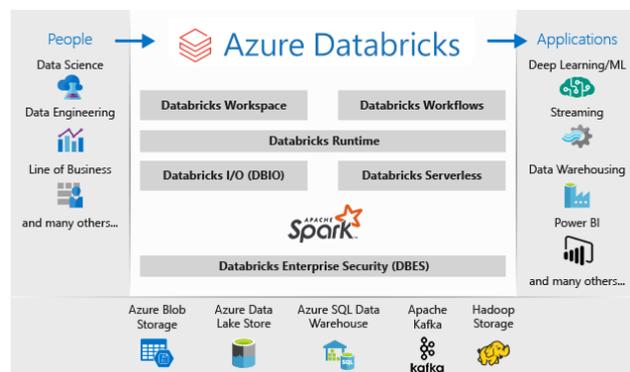


Fig. 2. Azure Databricks architecture

SQL Analytics in Azure Databricks provides a simple platform for data analysis in case of SQL queries against a data lake, supports a few types of visualizations needed results for special contexts, and using of dashboards.

The Azure Databricks Workspace provides an interactive workspace for collaboration between different specialists. In the big data pipeline, this data is received in Azure through Azure Data Factory as batches, or streamed in near real-time. This data ends up in a data lake for long-term storage in Azure Blob storage or Azure Data Lake Storage. As part of



your analytics workflow, you can use Azure Databricks to read data from a variety of data sources and get actionable insights using Spark.

The Azure Databricks workspace is a high-performance platform based on Apache Spark. The Azure Databricks workspace integrates with Azure, providing an interactive workspace, easy setup, and simplified workflows.

In the big data pipeline, this data (raw or structured) is received in Azure via the Azure Data Factory in packets or transmitted in near real-time streaming via Apache Kafka, the Event Hub, or the Internet of Things Center. This data enters a data lake for long-term storage in the Azure or Azure Data Lake Storage BLOB.

As part of the analytics workflow, you can use Azure Databricks to read data from multiple data sources, such as Azure BLOB, Azure Data Lake Storage, Azure Cosmos DB, or SQL Azure Data Storage, and retrieve useful statistics from them using Spark (Fig. 3).

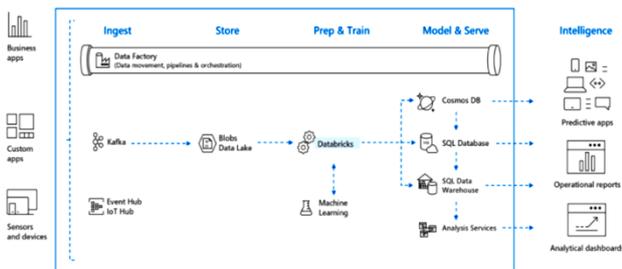


Fig. 3. Multiple data sources in Azure Databricks

The central component of **Amazon EMR** is the cluster. A cluster is a collection of Amazon Elastic Compute Cloud (Amazon EC2) instances. Each instance in the cluster is a node that has a role within the cluster, connected with the node type. Amazon EMR also installs different software components on each node type, giving each node a role in a distributed application alike Apache Hadoop (Fig. 4).

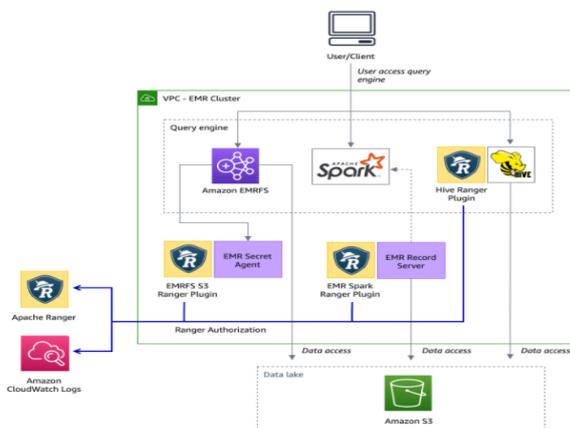


Fig. 4. Architecture of Amazon EMR

When you run a cluster on Amazon EMR, you have several options as to how you specify the work that needs to be done. When you launch your cluster, you choose the applications to install for your data processing needs. To process data in your Amazon EMR cluster, you can submit jobs or queries directly to installed software, or you can run steps in the cluster. A failure during the cluster lifecycle causes Amazon EMR to loss all data of it's instances in this cluster.

There are some benefits to using Amazon EMR: cost savings; AWS integration; deployment; scalability and flexibility; reliability; security; monitoring; management interfaces. But in this issue we are investigating to architecture's specific for different solutions first of all. Amazon EMR integrates with other AWS services to provide capabilities and functionality related to your cluster. Several examples of this integration: Amazon EC2 for the instances that comprise the nodes in the cluster; Amazon Virtual Private Cloud (Amazon VPC) to configure the virtual network in which you launch your instances; Amazon S3 to store input and output data; Amazon CloudWatch to monitor cluster performance and configure alarms; AWS Identity and Access Management (IAM) to configure permissions; AWS CloudTrail to audit requests made to the service; AWS Data Pipeline to schedule and start your clusters; AWS Lake Formation to discover, catalog, and secure data in an Amazon S3 data lake [14].

Azure HDInsight is a cloud distribution of Hadoop components. Azure HDInsight makes it easy, fast, and cost-effective to process massive amounts of data. HDInsight includes the most popular open-source Apache frameworks, such as: Hadoop, Spark, Hive with LLAP, Kafka, Storm, HBase and R (Fig. 5).

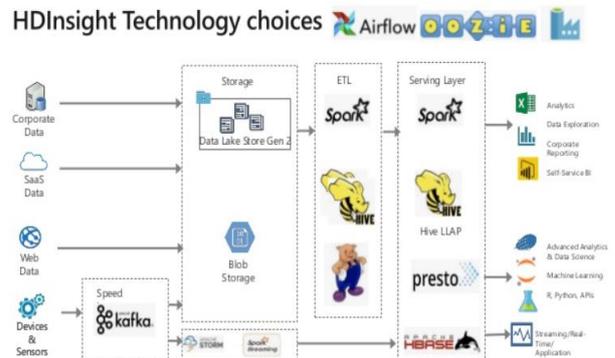


Fig. 5. Azure HDInsight Technology

Azure HDInsight advantages over on-premises Hadoop:

- Automated cluster creation requires minimal setup and configuration. Automation can be used for on-demand clusters.



- There's no need to worry about the physical hardware or infrastructure with an HDInsight cluster. Just specify the configuration of the cluster, and Azure sets it up.

- Azure takes care of data redistribution and workload rebalancing without interrupting data processing jobs.

- HDInsight is available in more regions than any other big data analytics offering.

- HDInsight enables you to protect your enterprise data assets with Azure Virtual Network, encryption, and integration with Azure Active Directory.

- A typical on-premises Hadoop setup uses a single cluster that serves many purposes. With Azure HDInsight, workload-specific clusters can be created. Creating clusters for specific workloads removes the complexity of maintaining a single cluster with growing complexity.

- You can use various tools for Hadoop and Spark in your preferred development environment.

- HDInsight clusters can be extended with installed components and can also be integrated with the other big data solutions by using one-click deployments from the Azure Market place.

- Azure HDInsight integrates with Azure Monitor logs to provide a single interface with which you can monitor all your clusters.

- HDInsight can easily be integrated with other popular Azure services such as the following: Data Factory (ADF), Blob Storage, Data Lake Storage Gen2, Cosmos DB, SQL Database, Analysis Services

- HDInsight constantly checks the infrastructure and open-source components using its own monitoring infrastructure. It also automatically recovers critical failures such as unavailability of open-source components and nodes. Alerts are triggered in Ambari if any OSS component is failed.

Snowflake Elastic Data Warehouse is an elastic system integrated with cloud storage on the Amazon Web Services platform, which can be expanded as needed as data, load, etc. increase

Snowflake provides the customer with a Data Warehouse as a Service. It is a high-performance column management system that supports standard SQL and meets ACID requirements. Data is accessed via Snowflake Web UI, Snowflake Client command-line interface, as well as ODBC and JDBC. The system consists of three components: Database Storage, Processing and Cloud Services.

The storage layer is responsible for the secure, secure and resilient storage provided by S3. Data is stored in S3, customers do not have direct access to it.

To upload data to Snowflake, special S3 buckets (staging area) are created, where you need to put files, from which you can then upload data using Snowflake SQL. When loaded, the data is compressed (gzip) and converted to column format. Indexes in Snowflake are not provided. Data distribution is carried out automatically on the basis of statistics of their use. There is no partitioning. (Fig. 6).

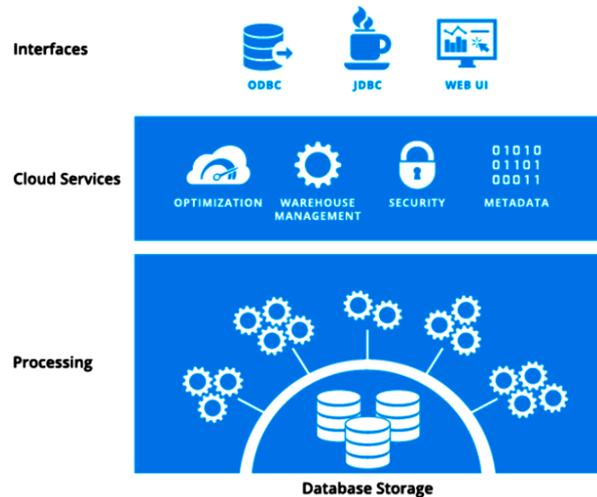


Fig. 6. Snowflake layers

Data is accessed through a data processing layer, a set of virtual servers that can access S3 files. A virtual cluster can consist of 1-16 virtual servers (EC2). All servers in the virtual cluster are the same and equivalent, the client does not have direct access to them. From the client's point of view, a virtual cluster is a whole. EC2 servers can be of two types: Standard and Enterprise. With Snowflake, it takes 1 to 5 minutes to create a new virtual cluster or expand an existing one. Data loss is not critical, in case of failure EC2 is simply re-created. SSD virtual servers are used as a cluster cache. Placing data in the cache speeds up queries up to 10 times. You can create multiple virtual clusters that will access the same data at the same time. This allows you to distribute the load.

Snowflake cloud services are used for data management. Using the Snowflake UI, the client creates databases, users and roles. Metadata is used at the data processing level (determining access rights, compiling queries, etc.). You can manage clusters: set the default cluster, specify the database to run, start the cluster on a schedule or when prompted, stop the cluster, change the number of servers. Snowflake provides installation, setup and maintenance. All you have to do is go to the site, create the tables, download the data, and run the query.



Modern systems and software stacks, (for example Apache Hadoop, for big data analytics are still complex to operate and far from perfect. As a result many organizations struggle to keep up with operating, optimizing and making their data infrastructure work to serve their data processing needs.

The complete data infrastructure solution includes many components among them:

Data Collection Service for both real time and bulk upload of data from different data sources such as applications, databases, web crawls etc.

Batch Computation Service such as Hadoop/Hive to process this data and transform it from data to information.

Real Time Computation Service to generate real time results on data streams and data captures for time sensitive and actionable reporting and monitoring.

AdHoc Query Service to answer one of queries sometimes exactly and other times approximately in a short amount of time.

Tools and Frameworks for job dependencies, data and query discovery, SLA and monitoring etc.

Qubole aims to provide all of the above components in the cloud. Qubole take care of optimizing, operating and evolving the data infrastructure for clients.

All data stored in S3. Adhoc Query and Batch Computation Service in the Cloud provides Apache Hive and Apache Hadoop as a service with close integration with Apache Oozie.

Hortonworks Data Platform (HDP) is open source Hadoop distribution that is based on a centralized architecture (Fig. 7).



Fig. 7. Architecture Hortonworks Data Platform

HDP addresses a range of data-at-rest use cases, powers real-time customer applications, and delivers robust analytics that accelerate decision making and innovation.

The following sections describe the HDP components:

- Data management
- Data access
- HDP Operations
- Security and governance

Data management. The foundational components of HDP are Yet Another Resource Negotiator (YARN) and Hadoop Distributed File System (HDFS). While HDFS provides the scalable, fault-tolerant, cost-efficient storage for a Hadoop-powered Big Data lake, YARN provides the centralized architecture that enables organizations to process multiple workloads simultaneously.

YARN also provides the resource management and pluggable architecture to enable a wide variety of data access methods.

Data access. With YARN at its architectural center, HDP provides a range of processing engines that allow users to interact simultaneously with data in multiple ways. YARN enables a range of access methods to coexist in the same cluster against shared datasets to avoid unnecessary and costly data silos. HDP enables multiple data processing engines that range from interactive SQL, real-time streaming, data science, and batch processing. These processing engines leverage data that is stored in a single platform, unlocking an entirely new approach to analytics.

HDP Operations. HDP Operations enables IT organizations to bring Hadoop online quickly by taking the guesswork out of manual processes and replacing them with automated preconfigured best practices, guided configurations, and full operation control. Because of the rapid emergence of Hadoop, many users lack an optimal way to provision and operate the environment, leading them to waste time on inefficient troubleshooting, monitoring, and configuration. HDP Operations makes it easy to operate distributed multi-user, multi-tenant, and multiple data access engines and helps manage HDP clusters at scale through an integrated web UI or single pane of glass. Apache Ambari is an open source management platform for provisioning, managing, monitoring, and securing Hadoop clusters. Ambari removes the manual and often error-prone tasks that are associated with operating Hadoop. It also provides the necessary integration points to fit seamlessly into the enterprise and enables the IT operator to focus on delivering world-class service and support for their HDP consumers.

Security and governance. Hortonworks created Data Governance Initiative (DGI), a consortium of cross-industry leaders, to address the need for an open source governance solution to manage data classification, lineage, security, and data life cycle management. Apache Atlas, created as part of the DGI, empowers organizations to apply consistent data classification across the data ecosystem. Apache Ranger provides centralized security administration for Hadoop.



CONCLUSIONS

HDFS is designed to be deployed on low-cost hardware. It provides high throughput access to application data. HDFS has been designed to be easily portable from one platform to another. It has a master/slave architecture.

The primary objective of HDFS is to store data reliably even in the presence of failures.

Simplified architecture Databricks allows traditional analytics and data science to co-exist in the same system. The complete vision of lakehouse architecture deliver 9x better performance than traditional cloud data warehouses.

The central component of Amazon EMR is the cluster. A cluster is a collection of Amazon EC2 instances. There are some benefits to using Amazon EMR? among them cost savings; AWS integration; deployment; scalability and flexibility; reliability; security; monitoring; management interfaces etc.

Azure HDInsight is a cloud distribution of Hadoop components. It makes it easy, fast, and cost-effective to process massive amounts of data.

The technology Snowflake is interesting, the ability to change the amount of resources on the fly looks very attractive. Cloud storage can be a decent option for new projects with limited data. But it's problem to transfer of all resources to this cloud.

Nowadays one of simplest and secure data lake platform for machine learning, streaming, and ad-hoc analytics was proposed by an Idera Inc. company and named Qubole.

HDP is the industry's only true secure, enterprise-ready open source Apache Hadoop distribution based on a centralized architecture (YARN). HDP includes a versatile range of processing engines that empower users to interact with the same data in multiple ways, at the same time. HDP extends data access and management with powerful tools for data governance and integration. Critical features for authentication, authorization, accountability and data protection are in place to help secure HDP across these key requirements. The users can integrate and extend their current security solutions.

REFERENCES

- [1] (2021) Mlitz K. Forecast revenue big data market worldwide 2011–2027 [Online]. Available: <https://www.statista.com/statistics/254266/global-big-data-market-forecast/>
- [2] Hajirahimova M. Sh., Aliyeva A. S. "Big Data initiatives of developed countries", *Problems of information society*, №1, pp. 10–15, 2017.
- [3] (2020) Patrisio A. Top Big Data Companies. [Online]. Available: <https://www.datamation.com/big-data/big-data-companies/>
- [4] (2021) Закон України про захист персональних даних [Online]. Available: <https://zakon.rada.gov.ua/laws/show/2297-17#Text>
- [5] Bradlow E. T., Gangwar M., Kopalle P. & Voleti S. "The Role of Big Data and Predictive Analytics in Retailing", *Journal of Retailing*, 93(1), pp. 79–95, 2017.
- [6] Chen S.-H., & Yu T. "Big Data in Computational Social Sciences and Humanities: An Introduction". *Big Data in Computational Social Science and Humanities*, pp. 1–25, 2018.
- [7] Fernando Y., Chidambaram R. R. M. & Wahyuni-TD I. S. "The impact of Big Data analytics and data security practices on service supply chain performance". *Benchmarking: An International Journal*, 25(9), pp. 4009–4034, 2018.
- [8] Gnizy I. "Big data and its strategic path to value in international firms". *International Marketing Review*, 36(3), pp. 318–341, 2019.
- [9] Harrison-Walker L. J. & Neeley S. E. "Customer Relationship Building on the Internet in B2B Marketing: A Proposed Typology". *Journal of Marketing Theory and Practice*, 12(1), pp. 19–35, 2004.
- [10] March Hofacker C. F., Malthouse E. C., & Sultan F. "Big Data and consumer behavior: Imminent opportunities". *Journal of Consumer Marketing*, 33(3), pp. 311–330, 2016.
- [11] (2017) Big data text analytics: An enabler of knowledge management. [Online]. Available: <https://doi.org/10.1108/JKM-06-2015-0238>
- [12] Kitchens B., Dobolyi D., Li J. & Abbasi A. "Advanced Customer Analytics: Strategic Value Through Integration of Relationship-Oriented Big Data". *Journal of Management Information Systems*, 35(2), pp. 540–574, 2018.
- [13] Liu C., Yang C., Zhang X. & Chen J. "External integrity verification for outsourced big data in cloud and IoT: A big picture". *Future Generation Computer Systems*, 49, pp. 58–67, 2015.
- [14] Liu X., Singh P. V. & Srinivasan K. "A Structured Analysis of Unstructured Big Data by Leveraging Cloud Computing". *Marketing Science*, 35(3), pp. 363–388, 2016.
- [15] Mawed M. & Aal-Hajj A. "Using big data to improve the performance management: A case study from the UAE. *FM industry. Facilities*, 35(13–14, SI), pp. 746–765, 2017.
- [16] Moorthy J., Lahiri R., Biswas N., Sanyal D., Ranjan J., Nanath K., & Ghosh P. "Big Data: Prospects and Challenges". *Vikalpa*, 40(1), pp. 74–96, 2015.
- [17] Salehan M. & Kim D. J. "Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics". *Decision Support Systems*, 81, pp. 30–40, 2016.
- [18] Sanders N. R. "How to Use Big Data to Drive Your Supply Chain". *California Management Review*, 58(3), pp. 26–48, 2016.
- [19] Szlezák N., Evers M., Wang J. & Pérez L. "The Role of Big Data and Advanced Analytics in Drug Discovery, Development, and Commercialization". *Clinical Pharmacology & Therapeutics*, 95(5), 492–495, 2014.
- [20] Talón-Ballesteros P., González-Serrano L., Soguero-Ruiz C., Muñoz-Romero S. & Rojo-Álvarez J. L. "Using big data from Customer Relationship Management information systems to determine the client profile in the hotel sector". *Tourism Management*, 68, pp. 187–197, 2018.
- [21] Tan K. H. & Zhan Y. "Improving new product development using big data: A case study of an electronics company". *R&D Management*, 47(4), pp. 570–582, 2017.

Стаття надійшла до редколегії

04.09.2021



Архітектура сучасних платформ для аналітики великих даних

Великі дані – це один із сучасних інструментів, який має великий вплив на світову індустрію, а також відіграє важливу роль у визначенні способів, якими підприємства й організації формують свою стратегію та політику. Проте кількість проведених наукових досліджень щодо прогнозування на основі великих даних є обмеженою через труднощі зі збиранням, обробленням та моделюванням неструктурованих даних, які зазвичай є конфіденційними. У статті великі дані розглядаються в контексті екосистеми для майбутнього прогнозування під час прийняття бізнес-рішень. Для однієї організації може бути важко володіти всіма необхідними даними, щоб розробити правильну стратегію. Ось чому різні організації створюватимуть та експлуатуватимуть власні аналітичні екосистеми або підключатимуться до існуючих. Екосистема аналітики, що складається із симбіозу даних, додатків, платформ, партнерств і сторонніх постачальників послуг, дозволяє організаціям бути більш гнучкими й адаптуватися до мінливих вимог сьогодення. Організації, які беруть участь в аналітичних екосистемах, можуть досліджувати, вчитися та впливати не лише на свої власні бізнес-процеси, а й на бізнес-процеси своїх партнерів. У цій публікації представлено архітектури популярних платформ для прогнозування на основі великих даних.

Ключові слова: великі дані, неструктуровані дані, платформи для аналізу даних, екосистеми даних, середовище великих даних.



Людмила Зубик. Закінчила факультет кібернетики Київського національного університету імені Тараса Шевченка, кандидат технічних наук, доцент. Працює на посаді доцента кафедри програмних систем і технологій факультету інформаційних технологій Київського національного університету імені Тараса Шевченка, Україна.

Сфера наукових інтересів: педагогіка вищої школи, вебтехнології, штучний інтелект.

Liudmyla Zubyk. Graduated from the Cybernetics Faculty of Taras Shevchenko National University of Kyiv, PhD, Associate Professor. She works as an associate professor at the Software Systems and Technologies Department, Faculty of Information Technologies, Taras Shevchenko National University of Kyiv, Ukraine.

Research interests: high school pedagogy, web technologies, artificial intelligence.



Ярослав Зубик. Закінчив факультет кібернетики Київського національного університету імені Тараса Шевченка, старший викладач. Працює на кафедрі комп'ютерних наук і прикладної математики факультету кібернетики та комп'ютерних наук Національного університету водного господарства та природокористування, Рівне, Україна.

Сфера наукових інтересів: проблеми оптимізації, аналіз даних.

Yaroslav Zubyk. Graduated from the Cybernetics Faculty of Taras Shevchenko National University of Kyiv, Senior Lecturer. He works at the Department of Computer Sciences and Apply Mathematics, Institute of Automatics, Cybernetics and Computer Engineering, National University of Water and Environmental Engineering, Rivne, Ukraine.

Research interests: optimization problems, data analytics.



ЦИФРОВІ ТЕХНОЛОГІЇ УПРАВЛІННЯ ІТ-ПРОЄКТАМИ



УДК 005.8

DOI: <https://doi.org/10.17721/AIT.2021.1.10>

В. В. Морозов, orcid.org/0000-0001-7946-0832,
М. В. Проскурін, orcid.org/0000-0002-6601-3133,
Київський національний університет імені Тараса Шевченка, Київ, Україна

Моделі проактивного управління змінами у проєктах створення та розвитку розподілених інформаційних систем

Розглянуто підходи до формування моделей проактивного управління проєктами створення та розвитку розподілених інформаційних систем. Досліджено особливості розподілених ІТ-проєктів як складної адаптивної системи, що розвивається за численних взаємопов'язаних обмежень. Зроблено висновок, що у цьому випадку самі проєкти носять розподілений характер. Проаналізовано основні тренди розвитку таких систем, виявлено основні закономірності та проблеми. Зосереджено увагу на розгляді базових складових компонент таких систем і запропоновано формалізовані моделі для опису їхньої взаємодії. Приділено увагу формуванню та реалізації проєктів щодо створення складних систем на базі розподілених інформаційних систем, де основні особливості полягають у дії значної кількості складних впливів із боку турбулентного зовнішнього оточення. Досліджено питання реакцій проєкту на впливи цих змін. Для вирішення питань з управління такими складними проєктами запропоновано застосування проактивного підходу. Досліджено наслідки змін у процесах створення вказаних систем. Побудовано математичну модель мінімізації впливів змін на елементи проєкту під час управління складними проєктами та запропоновано підхід до управління конфігурацією таких проєктів, що в цілому дозволяє побудувати схему процесу ефективного управління проєктами.

Ключові слова: проактивне управління, розподілені інформаційні системи, ІТ-проєкти, управління проєктами, конфігурація проєкту, впливи, макротренди.

Для цитування (for citation): В. В. Морозов, М. В. Проскурін. "Моделі проактивного управління змінами у проєктах створення та розвитку розподілених інформаційних систем," *Сучасні інформаційні технології*, vol. 1, p. 76–85, 2021.

ВСТУП

Сучасні тенденції розвитку хмарних технологій передбачають значні зусилля розробників щодо створення розподілених інформаційних систем (РІС). У свою чергу, розроблення методологій і систем управління проєктами дає можливість використовувати сучасні технології управління для забезпечення швидкого й ефективного створення таких РІС. Застосування проактивного підходу дозволяє будувати інтегровані системи управління шляхом створення складних ІТ-продуктів, що користуються експоненціально зростаючим попитом.

Специфіка управління розвитком РІС свідчить про створення розподілених ІТ-проєктів, включаючи створення та розвиток розподілених інформаційних систем. Нині це означає, що буде розв'язано низку досить складних проблем. Складність цих проблем безпосередньо залежить від складності проєкту та великої кількості скла-

дних поперечних впливів і взаємодій усередині та поза проєктом (середовище проєкту). Водночас проєкти та середовище перебувають у стані постійних змін і, як наслідок, піддаються постійній самомодифікації.

Як зазначено в [1], схеми взаємодії, що використовуються у реалізації проєктів і функціонуванні ІТ-компаній, зображено на рис. 1 як макротенденції та їхній вплив на проєкти й організації.

Макротренди визначають, як розробляються проєкти в компаніях, включаючи великих ІТ-гравців, і з якими особливостями проєкту мають справу ці компанії. Крім того, ці макротренди допомагають створити глобальну картину та скласти прогнози щодо результативності проєкту – витрати, доходи, терміни тощо.

Існуючі системи управління проєктами зосереджено на розв'язанні задач лінійного призначення [2]. Насправді, продемонстровані системи не спрямовані на ефективне управління проєктом

© Морозов В. В., Проскурін М. В., 2021



у цілому, оскільки на нього впливає сукупність складних взаємозалежних подій і пов'язаних із ними реакцій.

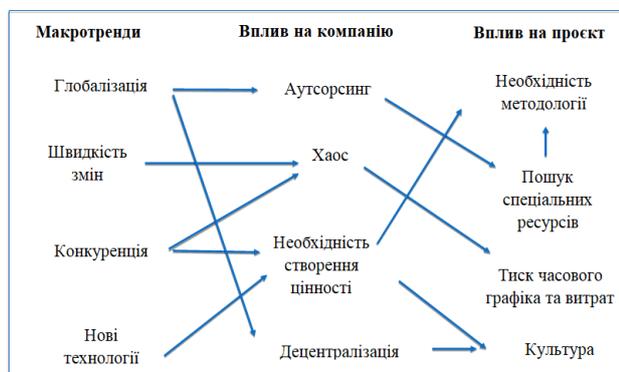


Рис. 1. Макротренди та впливи на створення РІС

Цей факт вимагає подальшого розроблення підходів до побудови систем управління для складних ІТ-проєктів, зокрема і використання проактивних інструментів. Ось чому важливо вивчати процеси попереджувального (випереджувального) управління для розроблення складних розподілених проєктів. Враховуючи зазначені тенденції та пов'язані з ними складні багатокритеріальні місії активного управління у створенні нових ІТ-продуктів, ми розглянемо подальший матеріал.

АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ

Використання ініціативних підходів до управління у проєктах, а також у розвитку організацій вивчали Н. Бушуєва, С. Бушуєв. У дослідженні [3] запропоновано модель “маріонетки”; автор дослідження зосереджений на технічних, управлінських і ділових аспектах як орієнтирах для збалансованого активного управління. Ці аспекти є сферами індукованої невизначеності, що великою мірою визначає моделі, які застосовують у проєктах.

Управління проєктами, засноване на проактивному підході, також вивчалось у роботах [4–7].

Автори, що розглядають проблематику та перспективи розвитку розподілених систем [8, 9], зазначають певну динаміку та тенденції у проєктах цього типу. Останнє доводить справжню складність і неоднозначність процесу. Тенденції, продемонстровані в [10], також виявляють деякі складні місії та неоднозначність в управлінні створенням і розвитком РІС. Тому видається раціональним використовувати активні заходи управління.

Проте використання проактивного управління у створенні та розвитку РІС є недостатньо ви-

вченим. Організації, що займаються такими дослідженнями, найбільш чутливі до згаданих макротрендів; отже, на створення РІС зрештою впливають зазначені тенденції. Цей факт пояснюється посиленням розвитком і використанням інноваційних технологій, їхніми швидкими змінами, попитом на ринку, а також необхідністю адаптації до сильно турбулентного середовища.

МОДЕЛІ УПРАВЛІННЯ ДЛЯ СТВОРЕННЯ ТА РОЗВИТКУ РІС

Динаміка розвитку проєкту визначається багатьма складними впливами. Впливи можна розділити на декілька груп: ті, що походять із зовнішнього середовища, ті, що походять з внутрішнього середовища організації, управлінські впливи та реакції проєкту на всі ці події. Іншими словами, у нас є чотири системи (продукт, проєкт, організація та середовище), де всі компоненти взаємодіють між собою, і кожна ситуація унікальна [11–13].

Оскільки хмарні технології є невід’ємною частиною розподілених інформаційних систем, цілий комплекс питань, пов’язаних з їхнім розробленням і використанням, стає предметом створення та розвитку РІС. Тому проблеми синхронізації даних із корпоративними та хмарними службами, безпеки та захисту даних, надійності, масштабованості тощо є важливими для аналізу в межах передбачуваного управління проєктами цього типу [14].

З огляду на наведені приклади проєктів створення та розвитку РІС, ми можемо виокремити основні сфери впливу у використанні активного управління: учасники проєкту, користувачі, акціонери, фізичні та програмні компоненти системи, правила взаємодії, що застосовуються до системних компонентів, мережних послуг, а також хмарні технології [15]. Це показано на моделі РІС, рис. 2.

Розширення кола місій для створення сучасних інформаційних систем, збільшення їхньої складності, неоднозначності та посилення турбулентності у зовнішньому середовищі, великих обсягів інформації вимагає нових технічних рішень для забезпечення ефективної інтеграції різномірних інформаційних ресурсів, що визначаються такими вимогами: надійність, безпека, незалежність від типу операційної системи чи мережі зв’язку тощо [16]. Серед цих інструментів можуть бути розподільні інформаційні системи (РІС), які зараз активно використовуються для розв’язання вказаних проблем і представляються універсальним підходом для різних сфер діяльності [17–20].

Оскільки створення РІС пов'язане з ІТ-проектом і характеризується не тільки функціональною та конструктивною складністю, але й великою кількістю змін на етапах планування, а особливо на етапах реалізації проекту.

Використання хмарних технологій у розподілених інформаційних системах трансформує сприйняття продукту як послідовність створення та розвитку РІС [21].

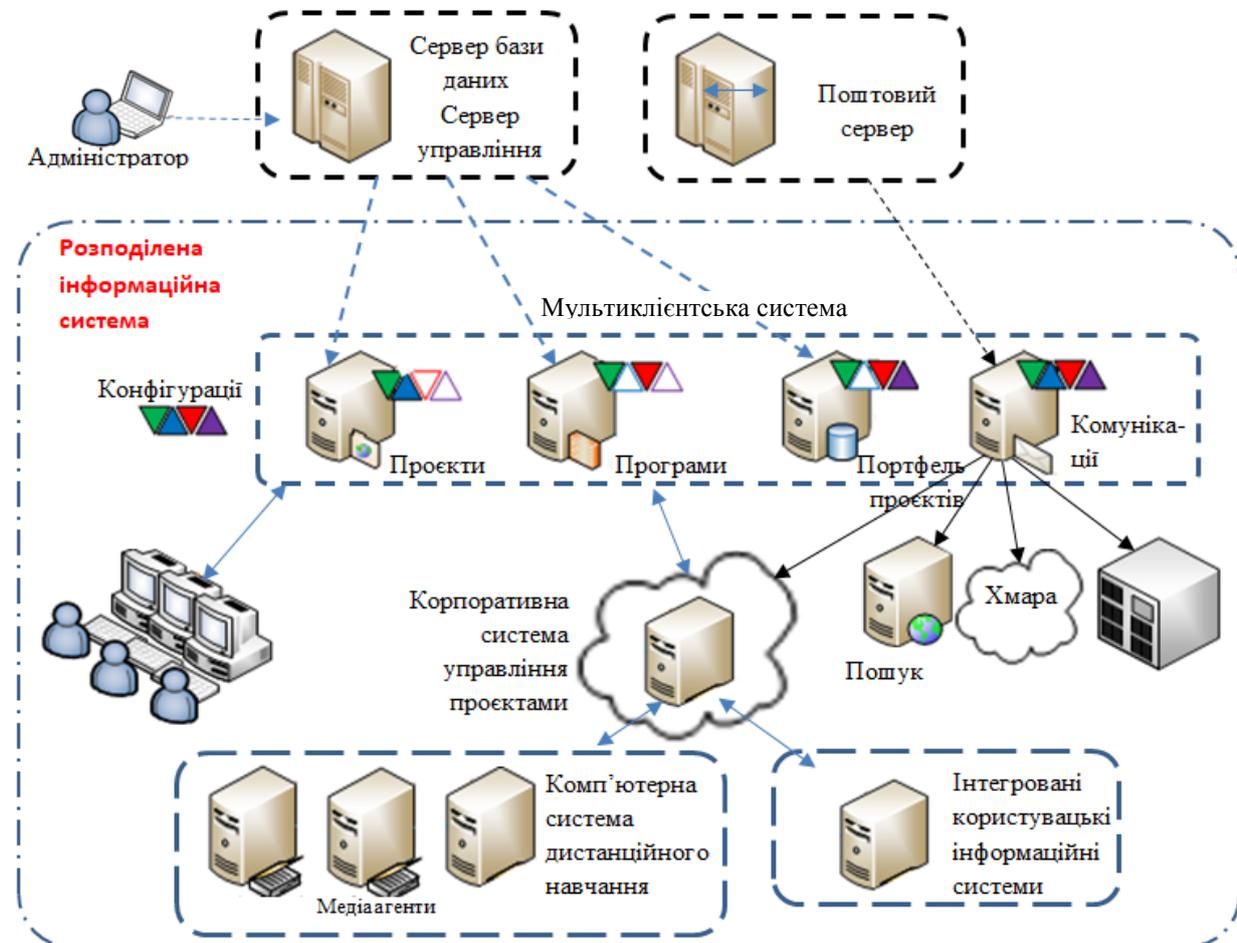


Рис. 2. Приклад розроблюваної розподіленої інформаційної системи управління проектами та програмами

Тому моделювання товару – це перший крок до активного управління проектами. Цей процес виконується на стадії концептуалізації проекту. Однак на ранніх стадіях концептуалізація товару недостатня для забезпечення ефективного проактивного управління моделюванням товару [22, 23]. Слід проводити систематичне оновлення затвердженої моделі, щоб зв'язати її параметри з фактичними умовами зовнішніх факторів.

Відмінною рисою активного управління є запобігання негативним подіям (інцидентам) шляхом контролю ключових показників, виявлення потенційних проблем на основі слабких сигналів і моделювання сценаріїв перебігу подій. Хоча справжня проблема полягає в управлінні великою кількістю параметрів [24], а також у контролі та прогнозуванні, що з бюджетного погляду є досить коштовним.

Оскільки для проектів цього типу часто потрібні знання з різних дисциплін і використання безлічі технологій, процес моніторингу й аналізу має включати всі аспекти проекту та компонентів продукту, а також враховувати взаємодію між компонентами й елементами. Окрім цього, ми не повинні нехтувати фізіологічним компонентом, визначеним учасниками проекту.

Ця місія ускладнюється ще й тому, що екологічні зміни реалізації проекту продовжують зростати, і відбулися зміни можна трактувати по-різному. Технічно це ускладнює або майже неможливо вирішити проблему.

Керуючи створенням та розвитком РІС, слід вжити заходи для запобігання виникненню інцидентів. Ці заходи включають:

- визначити оптимальний формат для конфігурації продукту. За допомогою цього процесу



ми забезпечуємо найвищу відповідність товару потребам бізнес-стратегічного плану [25];

- забезпечити інформаційну безпеку та запобігання інцидентам безпеки;
- підтримувати нормальне функціонування інфраструктури на експлуатаційному рівні завдяки попереджувальним випробуванням;
- керувати процесами тестування та випуску, щоб запобігти використанню компонентів продукту, які мають деякі проблеми;
- керувати всіма зацікавленими сторонами та контролювати конфігурацію продукції.

У свою чергу, ініціативне управління перебуває на вершині згаданих вище процесів, що об'єднують їхні цілі та місії. Водночас це інструмент, який дозволяє частково відійти від звичного формату управління змінами. Цей формат намагається привести елементи проекту до запланованого стану [4]. Хоча він не враховує об'єктивні причини виникнення змін і зміни зовнішніх факторів. Можливо, не варто намагатися повернутися до початкових параметрів проекту, оскільки ситуація суттєво змінилася, і ці показники вже не цікаві для клієнта і не будуть відповідати цілям проекту.

Крім того, деякі зміни можуть бути настільки масштабними, що їхній вплив на компоненти проекту або продукту незворотний. У цьому випадку всі зусилля щодо приведення проекту в початковий стан будуть абсолютно марними [26].

Тому всі зміни можна розділити на дві категорії:

- зміни, що приносять позитивну динаміку та перспективи;
- зміни, що негативно впливають на ефективність проекту.

Обидві зміни можна простежити за допомогою активного управління. Однак справжнє питання полягає в тому, як класифікувати передбачувані інциденти, коли вони виявляються на ранній стадії. Цей підхід дозволяє розвинути два типи реакції на потенційні зміни.

Перший підхід полягає в тому, щоб сприймати зміни як позитивний фактор і продумати шляхи їхнього використання для покращення результатів проекту. Як варіант, проект можна цілісно змінити і перетворити в нову концепцію, зберігаючи початкові цілі або трохи змінюючи їх, щоби збільшити вартість товару або поліпшити його важливість для бізнесу. Крім того, важливо бачити загальну картину майбутнього продукту та проекту, стимулювати креативність, чіткість сприйняття реальності, відкритість до змін і гнучкість у виборі реакції в нових умовах проекту. Це може бути "другим вітром" для проекту, його модернізації.

Другий пропонує дещо статичну реакцію та застосування класичних інструментів управління змінами. Але навіть у цьому випадку активні дії можуть принести позитивні результати, мінімізуючи шанси побачити негативний інцидент у повному обсязі.

Варто зазначити, що рівень негативного впливу на проект визначається скоріше сприйняттям та інтерпретацією реальності учасниками проекту, які оцінюють ці події. Хоча впевненість у можливості прогнозувати майбутні події часто є оманною, заснованою на хибному сприйнятті того, що прогноз може бути зроблений шляхом аналізу минулих подій.

Джерелами, які потенційно можуть спричинити зміни, є зони між складними адаптивними системами. Ці зони – середовище, на яке впливають сусідні системи; саме тут ми спостерігаємо посилену турбулентність. У вказаних зонах ми шукаємо слабкі сигнали для виявлення потенційних інцидентів, які матимуть позитивний вплив на реалізацію проекту.

Компоненти конфігурації ІТ-продукту [25, 27] пов'язані з визначенням елементів, їхніх параметрів і відношенням до розробленої інформаційної системи. Те саме стосується елементів проекту (рис. 3) та середовища проекту.

Перегляд впливу середовища проекту на його елементи та фактори успіху показує, що саме цей вплив (або ігнорування цього впливу) дуже часто є головною загрозою провалу проекту.

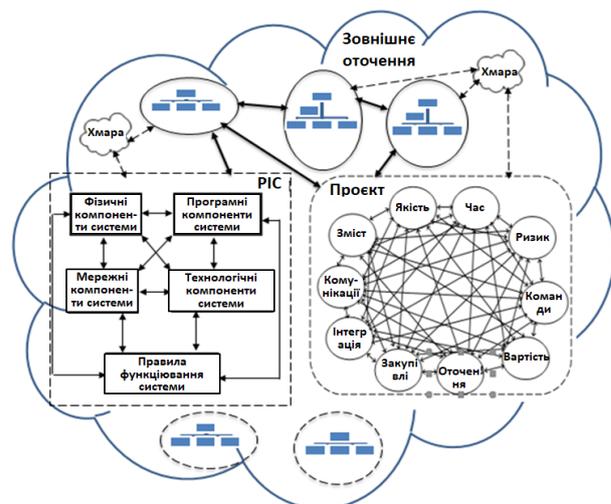


Рис. 3. Управління взаємодією елементів складних адаптивних систем

Вплив середовища проекту часто призводить до динамічних змін [28]. Вибірковий підрахунок цих змін веде до змін параметрів і характеристик практично всіх елементів в межах моделі "кону-



са". Причому питання проактивного впливу на функціонування розподілених інформаційних систем залишаються актуальними. Цю проблему можна розв'язати за допомогою активних підходів у проєктах із побудови таких систем.

Процеси створення продукту й управління IT-проєктами представлено як процеси в моделі "конуса".

$P = \{P^P, P^S\}$, де P^P – це сукупність процесів управління проєктами,

$P^P = \{p_1^P, p_2^P, \dots, p_i^P\}$, тут i – кількість процесів, пов'язаних з управлінням IT-проєктами; P^S – це сукупність процесів створення продукту,

$P^S = \{p_1^S, p_2^S, \dots, p_j^S\}$, де j – кількість процесів, пов'язаних із створенням IT-продукту.

Багато ресурсів у моделі "конуса" містять наявні та залучені (придбані) в межах проєкту матеріали, програмне забезпечення, робочу силу й інформаційні ресурси, які матеріалізуються в кінцевому продукті й утворюють нову бажану вартість у вигляді IT-проєкту продукту (PIC).

$R = \{R^M, R^H, R^P, R^I\}$, де R^M – це сукупність матеріальних ресурсів, задіяних у проєкті,

$R^M = \{r_1^M, r_2^M, \dots, r_l^M\}$, де l – кількість видів матеріальних ресурсів, необхідних у проєкті; R^H – сукупність людських ресурсів, що беруть участь у проєкті, $R^H = \{r_1^H, r_2^H, \dots, r_k^H\}$, тут

k – кількість видів людських ресурсів, залучених до проєкту; R^P – набір програмних ресурсів, що використовуються у проєкті,

$R^P = \{r_1^P, r_2^P, \dots, r_s^P\}$, де s – кількість видів програмних ресурсів, задіяних у проєкті; R^I – набір інформаційних ресурсів у проєкті t , $R^I = \{r_1^I, r_2^I, \dots, r_z^I\}$, тут z – кількість видів інформаційних ресурсів, задіяних у проєкті.

Зацікавлені сторони [6], що прямо впливають не тільки на функціональність майбутнього продукту проєкту, але й на успіх усього проєкту, суттєво позначаються на успіху IT-проєкту. Сюди можуть входити представники клієнта або замовника, користувачі IT-продуктів, постачальники, розробники, команди управління тощо.

$O^L = \{O^L, O^D\}$, де O^L – сукупність учасників проєкту (близьке оточення проєкту),

$O^L = \{o_1^L, o_2^L, \dots, o_f^L\}$, де f – це кількість учасників, які є частиною його близького оточення;

O^D – сукупність зацікавлених сторін (довгострокове середовище проєкту), $O^D = \{o_1^D, o_2^D, \dots, o_h^D\}$,

де h – це кількість зацікавлених сторін, які доцільні до довгострокового середовища проєкту.

Для IT-проєктів створення складних IT-продуктів характеризується сервісним компонентом, який вимагає стратегічного сервісу, служб розроблення, переходів, операцій і постійних вдосконалень [21]. $S = \{S^I, S^B, S^{US}\}$, де S^I – набір послуг, що підтримують IT-інфраструктуру,

$S^I = \{s_1^I, s_2^I, \dots, s_q^I\}$, де q – кількість служб підтримки IT-інфраструктури; S^B – набір послуг, що підтримують бізнес-додаток,

$S^B = \{s_1^B, s_2^B, \dots, s_w^B\}$, тут w – кількість служб для підтримки бізнес-додатків; S^{US} – набір служб, що надають підтримку користувачам,

$S^{US} = \{s_1^{US}, s_2^{US}, \dots, s_v^{US}\}$, де v – кількість служб підтримки користувачів.

Ми вже говорили про використання технологій для створення IT-продукту. Крім того, до цієї групи також входять технології для розроблення, управління, тестування, експлуатації та обслуговування продукту IT-проєкту. $Z = \{Z, Z^M, Z^I\}$,

де Z^C – набір технологій для розроблення та тестування продукту проєкту, $Z^C = \{z_1^C, z_2^C, \dots, z_c^C\}$,

тут c – кількість технологій, що використовуються для створення проєктного продукту; Z^M – набір технологій управління проєктами,

$Z^M = \{z_1^M, z_2^M, \dots, z_u^M\}$, де u – кількість технологій, що застосовуються для управління проєктом;

Z^I – сукупність технологій реалізації та підтримки проєктного продукту. $Z^I = \{z_1^I, z_2^I, \dots, z_y^I\}$,

тут y – кількість технологій, що використовуються для впровадження та підтримки проєктного продукту.

Компоненти конфігурації IT-продукту [19, 22] пов'язані з визначенням елементів, їхніх параметрів і відношенням до розробленої інформаційної системи. Те саме стосується елементів проєкту (рис. 1) та середовища проєкту.

$K = \{K^P, K^S, K^E, K^{DP}, K^S\}$, де K^P – набір параметрів проєкту, $K^P = \{k_1^P, k_2^P, \dots, k_r^P\}$, де

r – кількість параметрів проєкту; K^S – набір параметрів проєктного продукту,



$K^S = \{k_1^S, k_2^S, \dots, k_d^S\}$, де d – кількість параметрів продукту проекту; K^E – набір параметрів зовнішнього середовища проекту,

$K^E = \{k_1^E, k_2^E, \dots, k_\gamma^E\}$, де γ – кількість параметрів зовнішнього середовища проекту;

K^{DP} – набір вимог до проекту,

$K^{DP} = \{k_1^{DP}, k_2^{DP}, \dots, k_o^{DP}\}$, де o – це кількість вимог (умов, можливостей та обмежень), яким має відповідати проект; K^{DS} – набір вимог до продукту проекту, $K^{DS} = \{k_1^{DS}, k_2^{DS}, \dots, k_p^{DS}\}$, тут p – кількість вимог (умов, можливостей та обмежень), яким повинен відповідати проектний продукт.

Під потужністю продукту йдеться про сукупність $L = \{L^S, L^P, L^E\}$, де $L^S = \{l_1^S, l_2^S, \dots, l_\mu^S\}$ – набір значень пропускної здатності компонентів продукту, μ – кількість пропускних можливостей усіх компонентів продукту;

$L^P = \{l_1^P, l_2^P, \dots, l_\phi^P\}$ – кількість пропускних можливостей усіх компонентів продукту, ϕ – кількість пропускних можливостей усіх елементів проекту (процеси управління створенням продуктів і процеси управління проектом);

$L^E = \{l_1^E, l_2^E, \dots, l_x^E\}$ – сукупність значень впливу (ступенів впливу) зовнішнього середовища проекту, x – кількість значень впливу внутрішнього середовища проекту, а також впливів його зовнішнього середовища.

Доступність (під час створення продукту) визначається набором оцінок рівня виконання функцій і вимог (надійність, підтримка, справність, продуктивність, безпека).

$A = \{A^R, A^N, A^O, A^P, A^S\}$, де $A^R = \{a_1^R, a_2^R, \dots, a_\omega^R\}$ – сукупність оцінок, пов'язаних із надійністю, ω – кількість оцінок надійності;

$A^N = \{a_1^N, a_2^N, \dots, a_\sigma^N\}$ – набір оцінок рівня підтримки, σ – кількість оцінок можливостей підтримки; $A^O = \{a_1^O, a_2^O, \dots, a_\delta^O\}$ – сукупність оцінок справності, δ – кількість оцінок справності;

$A^P = \{a_1^P, a_2^P, \dots, a_\delta^P\}$ – набір оцінок ефективності, δ – кількість оцінок ефективності;

$A^S = \{a_1^S, a_2^S, \dots, a_\vartheta^S\}$ – набір оцінок безпеки,

ϑ – кількість оцінок безпеки.

З урахуванням цього, математичний опис запропонованої моделі «конус» можна представити таким чином:

$M = \{X, Y, H\}$, де $X = \{G, P, R, O, S, Z, K, L, A\}$ – набір вхідних параметрів моделі;

$Y = \{C_p, T_p, Q\}$ – сукупність вихідних параметрів, на основі якої ми визначимо ефективність процесів управління IT-проектами, де C_p – планова вартість створення елементів проекту, T_p – запланована тривалість життєвого циклу проекту (наведена) [6], O – якість проекту, що визначається якістю кінцевого продукту та якістю процесів реалізації проекту; H – набір каналів зв'язку між елементами моделі управління IT-проектами каналів зв'язку. $H = \{h_1, h_2, \dots, h_\varepsilon\}$, ε – кількість прямих зв'язків між усіма елементами моделі та $\bar{H} = \{\bar{h}_1, \bar{h}_2, \dots, \bar{h}_\varepsilon\}$, ε – кількість зворотних зв'язків між усіма елементами моделі.

У цьому випадку набір вихідних параметрів проектної моделі також може бути представлений у формі $X = \{x_i \mid i = 1, 2, \dots, N_1\}$, де N_1 – кількість областей знань моделі M_1 .

Тоді запланована вартість проекту буде виглядати так:

$$C_p = \sum_{i=1}^{N_1} \sum_{j_1=1}^{T_p} \sum_{i_2=1}^{\varepsilon} (C_1(x_i, t_{j_1}) + C_2(h_{i_2})), \quad (1)$$

за обмежень

$$\forall (x_i \in X) \cup (q_i \in Q) \exists t_{j_1} \in T_p, T_p \geq 0$$

$$\text{та } C_p \leq C_b, C_b \geq 0,$$

де C_b – бюджетна вартість проекту (інвестиції),

C_1 – функція витрат на створення елементів вхідних параметрів з $\{X\}$ та час $t_{j_1} \in T_p$,

C_2 – функція витрат каналів зв'язку між елементами моделі з $\{X\}$.

Перегляд впливу середовища проекту на його елементи та фактори успіху показує, що саме цей вплив (або ігнорування цього впливу) дуже часто є головною загрозою провалу проекту.

Вплив середовища проекту часто призводить до динамічних змін. Вибірковий облік цих змін веде до змін параметрів і характеристик практично всіх елементів у межах моделі "конуса". У цьому разі питання проактивного впливу на функціонування розподілених інформаційних

систем залишаються актуальними. Указану проблему можна розв'язати за допомогою активних підходів у проєктах із побудови таких систем.

Сучасні умови, в яких слід розробляти та реалізовувати складні ІТ-проєкти, характеризуються спадом, періодичним фінансуванням, плинністю основного персоналу, змінами в технологіях, зміною уподобань споживачів, зміною кон'юнктури ринку, споживачів, користувачів тощо [29, 30]. Усе це вимагає частих змін не тільки "конусної" моделі "бази", але і кластерів знань. Отже, вплив турбулентності навколишнього середовища необхідно вводити в модель "конуса". У процесі отримання моделі "конус" вплив цього середовища слід вивчати шляхом подальшого впливу змін на всі елементи та характеристики складних ІТ-проєктів. Для успішного завершення проєкту всіма цими змінами потрібно керувати. В іншому випадку часті зміни ведуть до хаотичних ненормальних невідповідностей елементів системи, що призводить до її виходу з ладу. На рис. 4 представлено запропоновану модель управління змінами для ІТ-проєктів.

Особливістю запропонованої моделі є перегляд взаємодії системи створення продукту та необхідної системи управління проєктами (змінami) в умовах складних пересічних впливів турбулентного середовища.

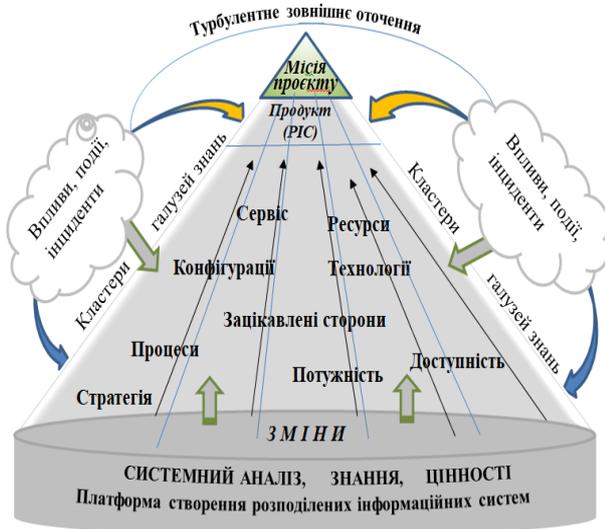


Рис. 4. Модель «конус» для дослідження впливів змін

Набори даних кожного із впливів генерують зміни в окремих компонентах моделі, у результаті чого змінюється процес створення розподілених інформаційних систем.

Ці зміни можуть призвести до проблем та інцидентів на різних етапах реалізації проєкту та продуктивності. Можливим рішенням уникнення

проблем або мінімізації їхнього впливу може бути раннє виявлення конкретних сигналів їхнього виникнення [31, 32].

Наявність турбулентного впливу середовища проєкту передбачає: додавання параметрів цього впливу до запропонованої моделі «конус» та реакції на неї як контрольних дій, що забезпечують стабілізацію моделі з виниклими відхиленнями.

Тоді модифікована модель проєкту матиме такий вигляд: $M_2 = \{X, Y, Q, I, U, V\}$, де $I = \{I^E, I^O\}$ – сукупність впливів на проєкт, I^E – сукупність впливів факторів навколишнього середовища, $I^E = \{i_1^E, i_2^E, \dots, i_e^E\}$, e – кількість можливих впливів зовнішнього середовища проєкту (політичний, економічний, соціальний, правовий, екологічний, технологічний аспекти); I^O – сукупність впливів зацікавлених сторін проєкту $I^O = \{i_1^O, i_2^O, \dots, i_b^O\}$, b – кількість можливих впливів проєкту з далекого та близького оточення (вторинні й основні зацікавлені сторони); U – сукупність станів ІТ-проєкту, $U = \{u_1, u_2, \dots, u_\beta\}$, β – кількість можливих станів моделі через вплив на довкілля та зацікавлені сторони проєкту; V – сукупність реакцій проєкту на зовнішні впливи, $V = \{v_1, v_2, \dots, v_a\}$, a – кількість керуючих дій, орієнтованих на стабілізацію моделі у разі відхилення її параметрів від заданих значень.

Беручи до уваги вплив зовнішнього середовища та зацікавлених сторін проєкту, що призводить до змін та відхилень від зазначених параметрів проєкту, можна визначити фактичну вартість проєкту після завершення (C_f) та фактичний час завершення проєкту (T_f):

$$T_f = T_p \pm (f_1(I) + f_2(U) + f_3(V)), \quad (2)$$

$$C_f = C_p \pm (C_2(I) + C_4(U) + C_5(V)), \quad (3)$$

де C_3, C_4, C_5 – фактичні витрати на внесення змін через сукупність впливів на проєкт, моніторинг набору станів ІТ-проєкту та набір виконуваних дій управління відповідно;

f_1, f_2, f_3 – функції для вимірювання часових інтервалів дії набору впливів на проєкт, моніторингу набору станів ІТ-проєкту та набору виконуваних дій управління відповідно.

У цьому випадку цільові функції моделі управління ІТ-проєктами можуть бути представлені таким чином:



$$C_f - C_p = \pm \Delta C \rightarrow \min, \quad (4)$$

$$T_f - T_p = \pm \Delta T \rightarrow \min, \quad (5)$$

де $\Delta C, \Delta T$ – фактичні відхилення у вартості та часі виконання проєкту з урахуванням змін, що додаються внаслідок сукупності впливів і впливів навколишнього середовища.

ВИСНОВКИ

З урахуванням специфіки хмарних технологій і неоднорідності середовища, де відбувається такий проєкт, аспекти створення та розвитку РІС, розглянуті в цій статті, вимагають їхнього подальшого вивчення. Виявлені особливості та запропоновані методи аналізу середовища, де відбувається взаємодія та самомодифікація систем, що беруть участь у створенні продукції, служать підставою для розроблення активних методів управління. Ці методи можна використовувати для управління концепцією продукту та проєкту, конфігурацією зацікавлених сторін, визначенням параметрів процесу для оптимального управління проєктами та успішного прогнозування. Крім того, перелічені аспекти можуть бути цікавими для вивчення впливу проактивного управління на створення та розвиток РІС як розподілених проєктів у портфелі розподілених проєктів організації.

Запропонований ініціативний підхід до управління змінами у проєктах розподілених інформаційних систем дозволив створити моделі, що відображають ключові елементи, які зазнають постійного впливу динамічних змін турбулентного середовища.

Для управління цими змінами слід використовувати запропоновану структуру процесів управління ІТ-проєктами на основі прийнятих.

Крім того, слід запропонувати модель обміну інформацією про активні компоненти управління. Такі процеси в цих системах слід додатково досліджувати та розробляти разом із побудовою відповідного математичного апарату.

Запропонована технологія управління змінами привела до побудови алгоритму, який визначає реакцію на їхній вплив на основі виявлення нових подій. Крім того, наслідки змін приймаються від вивчення запитів на зміни.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

[1] G. Shilo, N. Furmanova, “Statistically Oriented Tolerance Design with Correlation between Parameters of Components,” in *Proc. of the 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, Vol. 2, 21–23 September, 2017, Bucharest, pp. 1082–1088.

[2] (2016) A. Vajno, A. Kobiakov, V. Saraev, Anticipatory management. [Online]. Available: <https://dereksiz.org/uprejdajushee-upravlenie.html>

[3] Н. Бушуєва, *Моделі та методи проактивного управління програмами організаційного розвитку: монографія*, Київ: Науковий світ, 2007.

[4] V. Morozov, O. Mezentseva, “Trainable neural networks modelling for a forecasting of start-up product development”, in *Proc. of the IEEE 3rd International Conference on Data Stream Mining and Processing, DSMP*, 2020, pp. 55–60.

[5] A. Kolomiiets, V. Morozov, Investigation of optimization models in decisions making on integration of innovative projects, ser. *Advances in Intelligent Systems and Computing*, AISC. Berlin, Germany: Springer, vol. 1246, 2020, pp. 51–64.

[6] Р. Руденський, Ю. Г. Лисенко. *Очікуване управління складними економічними системами: моделі, методи, інструменти: Монографія*, Донецьк: Південний Схід, 2009.

[7] (2012) О. Коваленко, В. Курейчук. “Огляд проблем та аспектів хмарних обчислень та послуг”. [Online]. Available: <https://cyberleninka.ru/article/n/obzor-problem-i-sostoyaniy-oblachnyh-vychisleniy-i-servisov>

[8] Practice Standard for Project Configuration Management. Newtown Square, Pa: Project Management Institute, 2007.

[9] MIL-HDBK-61. Military Handbook. Configuration Management Guidance. USA. Department of Defense, 1997.

[10] Project Management Body of Knowledge (PMBOK), Project Management Institute, Four Campus Boulevard, Newton Square, PA 19073-3299, USA, 2018.

[11] (2020) Cloud Terminology – Key Definitions. [Online]. Available: <https://www.getfilecloud.com/cloud-terminology-glossary/>

[12] (2009) Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0167739X08001957>

[13] B. Furht, A. Escalante. (2010) Handbook of Cloud Computing. [Online]. Available: <https://link.springer.com/book/10.1007/978-1-4419-6524-0>

[14] O. Maimon, L. Rokach. (2018) Data Mining and Knowledge PICcovery Handbook. [Online]. Available: <http://www.bookmetrix.com/detail/book/ae1ad394-f821-4df2-9cc4-cbf8b93edf40>

[15] М.С. Косяков, Введение в распределенные вычисления. СПб: Государственный научно-исследовательский институт ИСТ, 2014.

[16] М. Алпатов, “Развитие распределенных технологий и систем”, *Перспективы науки и образования*, 2015, т. 2 (14).

[17] В.Н. Бурков, С.Д. Бушуев, *Управление ресурсами у розподілених проєктах та програмах: Монографія*, Миколаїв: Видавець Торубара В.В., 2015.

[18] A. Biloshchytskyi., A. Kuchansky, Yu. Andrashko, S. Biloshchytska, “A method for the identification of scientists' research areas based on a cluster analysis of scientific publications,” in *Eastern-European Journal of Enterprise Technologies*. – 2017. – No.5. – Vol. 2. – Issue 89. – pp. 4–10.

[19] Yu. Teslia, A. Khlevnyi, I. Khlevna, “Control of informational Impacts on project management”, in *Proc. of the 1th IEEE International Conference on Data Stream Mining & Processing*, 23–27 August, 2016, Lviv, Ukraine.

[20] І.В. Чумаченко, В.В. Морозов, *Управление проєктами: планування проєктних дій. Підручник*. Київ: Університет економіки та права «КРОК», 2011.



- [21] V. Morozov, O. Mezentseva, M. Proskurin, “Trainable neural networks modelling for a forecasting of start-up product development”, in Proc. of the 2020 IEEE 3rd International Conference on Data Stream Mining and Processing, 2020, pp. 55–60.
- [22] (2017) Proactive Project Management. [Online]. Available: <http://www.itexpert.ru/rus/ITEMS/200810062247/>
- [23] Н. Бушуєва, Моделі та методи проактивного управління програмами організаційного розвитку. Навчальний посібник. Київ: Науковий світ, 2007.
- [24] (2011) Что такое ITIL/ITSM. [Online]. Available: <https://itsm365.ru/blog/chto-takoe-itilitsm/>
- [25] V. Morozov, A. Kolomiets, “Investigation of optimization models in decisions making on integration of innovative projects”, in Advances in Intelligent Systems and Computing, 2021, 1246 AISC, pp. 51–64.
- [26] S. Warrilow. (2010) Change management: the horror of it all, Project Smart. [Online]. Available: <https://www.projectsmart.co.uk/change-management-the-horror-of-it-all.php>.
- [27] Р. Тернер, *Руководство по проектно-ориентированному управлению*, пер. с англ., Москва: ИД Гребенникова, 2007.
- [28] L.T. Wilson, C.A. Snyder, “Knowledge Management and IT: How Are They Related?”, IT Pro, vol. 1, no. 2, 1999.
- [29] Larry Stevens, “At a Moment’s Notice: Final Mile Introduces Knowledge Management to a Project Already Underway”, Knowledge Management, 2001.
- [30] Komashinsky, Smirnov DA, *Neural networks and their use in control and communication systems*. Hotline-Telecom, 2003.
- [31] N. Yehorchenkova, Iu. Teslia, O. Yehorchenkov, “Method of project and operational processes integration in the activities of project-oriented enterprises based on functional 4P-environment,” in Proc. CEUR Workshop, 2020, pp. 142–151.
- [32] Luntovskyy, “Up-to-date Paradigms for PICtributed Computing”, in Proc. of the 2nd IEEE International Conference on Advanced Information and Communication Technologies-2017 (AICT), Lviv Polytechnic National University, July 3–7, 2017, Lviv, Ukraine., pp. 113–120.

Стаття надійшла до редколегії

18.10.2021



Models of proactive change management in projects for the creation and development of distributed information systems

This article considers approaches to the formation of models of proactive management of projects for the creation and development of distributed information systems. Features of distributed IT projects as a complex adaptive system that develops under numerous interrelated constraints are considered. It is concluded that the projects themselves are distributed. The main trends in the development of such systems were analyzed, the main patterns and problems were identified. The focus is on the consideration of the basic components of such systems and formalized models are proposed to describe their interaction. Attention is paid to the formation and implementation of projects to create complex systems based on distributed information systems, where the main features are the action of a significant number of complex influences from the turbulent external environment. The issues of project reactions to the effects of these changes were investigated. To address the management of such complex projects, a proactive approach is proposed. The consequences of changes in the processes of creating such systems are studied. A mathematical model of minimizing the effects of changes on project elements in the management of complex projects is proposed and an approach to the configuration management of such projects is proposed, which in general allows to build a scheme of effective project management.

Keywords: proactive management, distributed information systems, IT projects, project management, project configuration, impacts, macro trends.



Віктор Морозов. Завідувач кафедри технологій управління факультету інформаційних технологій Київського національного університету імені Тараса Шевченка, канд. техн. наук, професор, Київ, Україна. Наукові інтереси концентруються на проблемах управління складними розподіленими ІТ-проектами з використанням хмарних технологій. Член IEEE.
Viktor Morozov. PhD, Professor, Head of Dept. of Technologies Management of Faculty of Informational Technologies of Taras Shevchenko National University of Kyiv. Kyiv, Ukraine. Current research interests include problems of complex distributional IT projects with using cloud technology. Member of IEEE.



Максим Прокурін. Аспірант кафедри технологій управління факультету інформаційних технологій Київського національного університету імені Тараса Шевченка. Київ, Україна. Наукові інтереси: розроблення та реалізація інноваційних проєктів у галузі ІТ із використанням інтелектуальних технологій.
Maksym Proskurin. Postgraduate student of Dept. of Technologies Management of Faculty of Informational Technologies of Taras Shevchenko National University of Kyiv. Kyiv, Ukraine. Current research interests include development and implementation of innovative projects in the IT field using intelligent technologies.



УДК 005.8

DOI: <https://doi.org/10.17721/AIT.2021.1.11>

О. Г. Тімінський, orcid.org/0000-0001-8265-6932,
А. С. Коломієць, orcid.org/0000-0003-4252-5975,
О. О. Мезенцева, orcid.org/0000-0002-8430-4022,

Київський національний університет імені Тараса Шевченка, Київ, Україна

Моделі управління проектом створення ІТ-компанії у галузі Data Science

Розглянуто підґрунтя виникнення проектів створення ІТ-компаній у галузі Data Science, обґрунтовано актуальність таких проектів. Доведено доцільність їхнього дослідження з використанням моделей і методів управління проектами. Проаналізовано літературні джерела у трьох напрямках – класичний проектний менеджмент, гнучкі підходи до управління проектами, ціннісно-орієнтоване управління. Доведено недостатню дослідженість описуваної тематики. Виокремлено групи моделей і методів, які описують відповідні галузі знань з управління проектами, що будуть необхідними і мінімально достатніми для розроблення наукового базису проекту створення ІТ-компанії у галузі Data Science. А саме: управління командою проекту, управління цінностями проекту, розроблення концепції проекту, управління змістом проекту, управління комунікаціями проекту, гнучкі інструменти управління проектом. У процесі аналізу кожної галузі виокремлено відповідні моделі й методи та проаналізовано їхню застосовність до досліджуваного проекту. Запропоновано модель вибору множини моделей і методів у межах визначених ключових галузей знань для досліджуваного проекту у вигляді зортки критеріїв. Проведено SWOT-аналіз запропонованого підходу, розглянуто його сильні та слабкі сторони, можливості й загрози, що пов'язані з ним. Сформульовано висновки до проведеного дослідження. Окреслено перспективи подальших досліджень у обраному напрямі.

Ключові слова: інформаційні технології, управління проектами, Data Science, ІТ-компанія.

Для цитування (for citation): О. Г. Тімінський, А. С. Коломієць, О. О. Мезенцева. “Моделі управління проектом створення ІТ-компанії у галузі Data Science,” *Сучасні інформаційні технології*, vol. 1, p. 86–94, 2021.

ВСТУП

Сучасний світ проектного менеджменту рухається в напрямку інтегративної конгломерації різногалузевих досягнень, агрегатором яких виступає ІТ-галузь. Тренд діджиталізації охоплює усе більше галузей економіки і сфер суспільного життя. У руслі цього тренду актуалізується науково-практичний напрям Data Science як уособлення реалізації когнітивно-гносеологічних цінностей людства щодо відтворення багатоаспектності людської істоти в речах іншої природи – єдиній технічній системі чи певному їх комплексу.

Науковий аналіз і синтез у цьому контексті може бути спрямований не тільки на розроблення моделей і методів, які безпосередньо реалізують надбання Data Science щодо збирання, оброблення, інтерпретації та використання великих обсягів даних про певний об'єкт або певну сутність об'єктивної реальності. Цікавим із наукового погляду, актуальним із практичного і значущим із погляду науково-прикладного забезпечення розвитку суб'єктів господарювання і промислового клас-

теру загалом, є напрям створення певного організаційного утворення, спрямуванням якого буде діяльність із формування моделей і методів Data Science в одній із численних його підгалузей.

ПОСТАНОВКА ЗАДАЧІ

Отже, з наведеного вище можна зробити висновки про те, що створення організації (компанії, підприємства), яка працювала б у галузі Data Science, є актуальним практичним, а розроблення моделей щодо її створення – актуальним науковим завданням.

Резонно було б припустити, що створення такої організації може розглядатися як проект, оскільки має всі потрібні відповідні атрибути – обмеженість у часі, унікальність, паралельність здійснення щодо основної будь-якої діяльності учасників відповідної команди.

Тобто науковим базисом розгляду діяльності щодо створення суб'єкта у галузі розроблення напрямку Data Science доцільно обрати управління проектами, що відоме своїми напрацю-

© Тімінський О. Г., Коломієць А. С., Мезенцева О. О., 2021



ваннями у контексті створення і розвитку організаційно-технічних систем та інституцій. Проаналізуємо ж такі напрацювання.

АНАЛІЗ ОСТАННІХ ДОСЯГНЕНЬ І ПУБЛІКАЦІЙ

Намагаючись досліджувати діяльність у проєкті створення ІТ-компанії у галузі Data Science, маємо проаналізувати підходи, розглянувши відповідну діяльність із декількох ракурсів.

Перший ракурс дозволяє позиціонувати таку проєктну діяльність за десятьма аспектами, що відповідають десяти галузям знань із проєктного менеджменту, як це визначає найбільш знаний і відомий у світі управління проєктами стандарт РМВОК (Guide to Project Management Body of Knowledge) [1] Американського інституту проєктного менеджменту PMI (Project Management Institute), який позиціонує себе як міжнародна асоціація. Підхід моделювання досліджуваного проєкту з погляду РМВОК вимагає пропонування моделей і методів за всіма такими галузями знань, а саме: щодо управління розкладом, вартістю, ресурсами, ризиками, змістом проєкту, управління якістю, закупівлями, комунікаціями, інтеграцією у проєкті, управління залученням стейкхолдерів у проєкт. Подібні підходи можна реалізувати, скориставшись напрацюваннями інших стандартів – Prince2 (сім галузей знань) [2], ISO 21500 (40 процесів) [3] тощо. Доповненням цього підходу можна вважати напрацювання Міжнародної асоціації управління проєктами IPMA, зокрема у частині стандарту ICB (Individual Competence Baseline) [4], що формулює вимоги до компетенцій фахівця з управління проєктами. Такі вимоги можуть бути використані у прив'язці до кожного процесу процесної моделі, яка охарактеризована вище, що збагатить відповідний управлінський інструментарій.

Другий ракурс розглядає управління ІТ-проєктами через реалізацію гнучкого підходу, або, як вона самопозиціонується, гнучкої методології управління проєктами Agile [5, 6]. Основний концепт, що відрізняє цю методологію від класичних стандартів – використання моделі життєвого циклу, в якій реалізується інкрементний підхід. Сутність такого підходу полягає у поділі другої фази життєвого циклу ІТ-проєкту на окремі проміжки часу (у фреймворку Scrum, що належить до гнучкої методології, такі проміжки називають спринтами), у межах яких створюється наступний інкремент ІТ-продукту, який потім розглядається командою проєкту спільно із замовником. За результатами такого розгляду

спільно формулюється завдання (беклог) наступного спринту. Кардинальність новизни такого підходу є дискусійною, оскільки реалізація ІТ-проєкту все одно проходить три водоспадні фази (як би не критикували апологети Agile відповідну модель життєвого циклу) – PreGame (підготовка проєкту, розроблення беклогу проєкту тощо), Game (безпосередньо спринти), PostGame (фіналізація впровадження ІТ-продукту у замовника, остаточні розрахунки тощо). Однак Agile почали застосовувати і поза межами ІТ-сфери, зокрема (але не обмежуючись) в такій, що щойно випущена, – сьомій редакції класичного стандарту РМВОК [7], пропонується гібридна (з урахуванням Agile) концепція управління проєктами.

Третій ракурс відображає сучасну тенденцію до ширшого, відносно класичного, погляду на управління загалом і управління проєктами зокрема, із включенням у стратегічні сутності проєктної діяльності поняття цінностей і оперування (управління) такими цінностями. Стандарт Японської асоціації управління проєктами P2M (Program and Project Management for enterprise innovation) постулює типову модель системи цінностей для використання у проєктному менеджменті [8]. Ця модель визначає п'ять типів цінностей – цінності активу, цінності інновацій, цінності володіння для зацікавлених сторін, цінності інтелектуального активу. Зрозуміло, що передостанній тип цінностей може бути категоризований відповідно до множини основних зацікавлених сторін окремого проєкту. Отже, розроблення моделей і методів для управління досліджуванним проєктом відповідно до ціннісного ракурсу полягатиме у запропонуванні як моделей самих цінностей, так і методів і методів управління ними в межах діяльності щодо реалізації проєкту.

Наукові дослідження, що присвячені структуротворенню суб'єктів ІТ-простору [9, 10], з одного боку, не відображають у пропонованих моделях і методах специфічних рис компаній у галузі Data Science. З іншого боку, у дослідженнях галузі Data Science [11, 12] недостатньо розглянуто діяльність (як було доведено вище – проєктна за своїм характером) щодо створення організацій для функціонування в галузі. А отже виникає методологічний розрив, на подолання частини якого спрямована ця публікація.

МЕТА ДОСЛІДЖЕННЯ

Метою нашого дослідження є таке: проаналізувати застосовність моделей і методів управління проєктами (у шести аспектах) до проєкту створення ІТ-компанії у галузі Data Science; за-



пропонувати модель вибору множини моделей і методів у межах визначених аспектів для досліджуваного проекту.

ОСНОВНИЙ МАТЕРІАЛ ДОСЛІДЖЕННЯ

Рухаючись у напрямку розроблення синтетичного підходу до створення моделей досліджуваного проекту, беручи до уваги концепт «конвергенції методологій» [13], виокремимо групи моделей і методів (які описують відповідні галузі знань з управління проектами), що будуть необхідними і мінімально достатніми для ґрунтації наукового базису проекту створення ІТ-компанії у галузі Data Science.

Управління командою проекту. Основною рушійною силою будь-якого проекту є команда його реалізації. А отже, саме цей аспект проектного менеджменту вимагає найбільш обґрунтованого наукового підходу. Проектний світ напружував достатньо розробок. З них вибрано найпридатніші для досліджуваного проекту, їх охарактеризовано у табл. 1.

Таблиця 1
МОДЕЛІ УПРАВЛІННЯ КОМАНДОЮ ПРОЄКТУ

№ з/п	Аналіз моделей		
	Тип моделей	Різновиди моделей	Застосовність до проекту
1	Стилі управління	Авторитарний, демократичний, ліберальний, демократура, комбінований, інший	Демократура або комбінований
2	Моделі команд	Ієрархія виконання, неієрархічна команда, креативна команда (генерація прийняття рішень), дослідницька команда, гнучка команда (Agile), команда стартап-проекту, інша	Креативна команда або команда стартап-проекту
3	Моделі мотивації	Матеріальна, кар'єрна, заохочувальна, мотивація досягнень, мотивація навчання, комбінована, інша	Мотивація досягнень
4	Моделі KPI	За вчасністю реалізації, за якістю, за мінімальністю вартості (собівартості), за швидкістю розв'язання конфліктів, за ступенем згуртованості команди, за швидкістю прийняття рішень, інша	За швидкістю розв'язання конфліктів і за ступенем згуртованості команди

Підбір моделей у межах кожного типу для досліджуваного проекту пропонується здійснювати за загальним принципом, що буде наведений нижче, для всіх описуваних галузей знань.

Управління цінностями проекту. Ціннісно-орієнтоване управління базується в проектному менеджменті, перш за все, на моделях і методах японського стандарту P2M [8]. Концептуально стандарт запроваджує новий підхід до суті проекту, визначивши його як зобов'язання створити цінність (цінності). Транспонуючи підхід, викладений у стандарті, до досліджуваного проекту і розширюючи його на операції над цінностями, наведемо перелік моделей, що будуть застосовними до нього (табл. 2).

Таблиця 2
МОДЕЛІ УПРАВЛІННЯ ЦІННОСТЯМИ У ПРОЄКТІ

№ з/п	Аналіз моделей		
	Моделі цінностей / управління цінностями	Різновиди цінностей	Застосовність до проекту
1	Цінність активу	Устаткування, виробництво (завод, підприємство), компанія, група компаній, наукова розробка, ІТ-платформа, набір інструментів для бізнесу (бізнес-модель) тощо	ІТ-платформа та набір інструментів для бізнесу (бізнес-модель)
2	Цінності інновацій	Технологічні інновації, ІТ-інновації, управлінські інновації, маркетингові інновації тощо	ІТ-інновації та маркетингові інновації
3	Цінності зацікавлених сторін	Цінності замовника, цінності команди, цінності учасників проекту, цінності споживача, цінності (інших) зацікавлених сторін тощо	Цінності команди та цінності потенційних споживачів
4	Цінності інтелектуального активу	Юридично оформлені ноу-хау (патент, інтелектуальні права), цінність сформованої команди, цінність створеної і відлагодженої (гнучкої) організації, цінність здатності до масштабування, цінність направлених процесів управління тощо	Цінність сформованої команди, цінність направлених процесів управління та цінність здатності до масштабування
5	Моделі створення цінностей	Креативні моделі, моделі копіювання, моделі бенчмаркінгу, моделі конвергенції тощо	Креативні моделі та моделі конвергенції
6	Моделі культивування цінностей	Використання харизматичного лідера, влада прикладу, заохочення носіїв цінностей тощо	Заохочення носіїв цінностей
7	Моделі моніторингу створення цінностей	Модель згортки критеріїв, модель Earned Value, модель кількісного оцінювання за відсотком досягнення тощо	Модель згортки критеріїв



Ціннісно-орієнтоване управління згенерувало значно більшу кількість моделей, які проте, з огляду на природну обмеженість обсягу цієї публікації, не увійшли до результатів аналізу – це і моделі профілювання місії, і моделі еталонного учасника команди, і моделі компетенцій команди тощо.

Розроблення концепції проєкту. Ініціалізація проєкту вимагає створення концепту – основи проєктної діяльності, в якій відображається особливість продукту проєкту, його «інакшість» відносно подібних продуктів, що є передумовою отримання конкурентних переваг після завершення проєкту і розпочинанню бізнес-фази життєвого циклу продукту, коли відбувається його комерціалізація.

Концепція проєкту відповідно до класичних стандартів складається, як правило, з декількох суттєвих елементів, моделі яких мають бути запропоновані командою проєкту. Розглянемо такі елементи (табл. 3).

Таблиця 3
Моделі створення концепції проєкту

№	Аналіз моделей		
	Елементи концепції	Моделі елементів	Застосовність до проєкту
1	Конфігурація продукту проєкту	Структура продукту (перелік елементів), характеристика елементів, функціональність продукту, опис інтерфейсів, вимоги до ІТ-засобів, платформи реалізації (якщо продукт – компанія, тоді до конфігурації входить юридична форма, модель оргструктури, моделі процесів тощо)	Усе перелічене
2	Конкурентна перевага	Моделі технологічного лідерства, моделі стратегічного лідерства, моделі PR, моделі зниження собівартості	Моделі технологічного лідерства, стратегічного лідерства та моделі PR
3	Маркетингова концепція	Портрет споживача (замовника), моделі маркетингу, моделі продажів, моделі підвищення іміджу компанії (продукту)	Усе перелічене
4	Концепція управління проєктом	Методологія управління проєктом	Класична або Agile або стартап
5	Концепція команди	Моделі компетенцій учасників команди, модель командної взаємодії (командотвірні концепти), моделі проведення нарад тощо	Усе перелічене

Загалом, моделі, що використовуються на етапі концептуального проєктування проєкту, мають більший спектр різновидів, однак для досліджуваного проєкту наведений перелік вважаємо мінімально достатнім.

Управління змістом проєкту. Узгоджена конфігурація продукту проєкту з особливостями, що потенційно містять конкурентні переваги, вимагає реалізації, яка відтворюватиме певну діяльність щодо створення такої конфігурації. Перелік указаних дій складає зміст проєкту, який потрібно визначити, контролювати та, за необхідності, змінювати, що є сутністю управління змістом проєкту.

Перелік моделей і методів, що можуть бути використані у цьому випадку, може підбиратися як зі стандартного кейсу, визначеного ustalеними міжнародними стандартами, так і створюватися під окремий тип проєктів. Орієнтовний перелік основних методів управління змістом наведено в табл. 4.

Таблиця 4
Методи управління змістом проєкту

№	Аналіз методів		
	Елементи управління змістом	Методи, що застосовуються в межах елементів	Застосовність до проєкту
1	Розроблення змісту	Метод аналогій, метод конвергенції, креативні методи (мозковий шторм, метод Дельфі тощо) та ін.	Креативні методи
2	Затвердження змісту	Метод послідовних узгоджень, метод сесій спільного розроблення (JAD Session) тощо	Метод сесій спільного розроблення (JAD Session)
3	Моніторинг виконання змісту	Методи управління якістю, метод Earned Value, календарно-сітьові методи тощо	Усі перелічені
4	Внесення змін до змісту	Методи управління якістю, метод сесій спільного розроблення (JAD Session) тощо	Усі перелічені

Окремо зазначимо, що зміст відіграє вирішальну роль в успіху проєкту створення ІТ-компанії у галузі Data Science (під успіхом тут розуміємо комерційну успішність компанії на ринку, її стійкість до змін зовнішнього середовища і гнучкість системи управління). Основним проблемним питанням змісту тут є визначення, у межах затвердженої концепції, такого змісту продукту самої компанії (тобто, ключового Data Science продукту, який буде реалізовувати створювана компанія, яка сама є продуктом



проєкту), котрий після завершення проєкту мав би комерційний успіх. Вибір такого Data Science продукту є складною творчою, дослідницькою, маркетинговою, у чомусь прогностичною задачею, розгляд якої перебуває поза межами цієї статті.

Управління комунікаціями проєкту. Інтеграція команди у напрямку успішного отримання узгодженого із замовником проєктного результату відбувається через комунікації, що є основою управління в широкому спектрі проєктної діяльності, охоплюючи всі його процеси – від ініціалізації, через планування, реалізацію, контроль і аж до завершення (як проєкту в цілому, так і кожної його галузі – управління розкладом, управління вартістю, управління якістю тощо). Розглянемо відповідні моделі (табл. 5).

Таблиця 5
Моделі управління комунікаціями у проєкті

№	Аналіз моделей		
	Елементи комунікацій	Моделі елементів	Застосовність до проєкту
1	Одиничні комунікації	Матриця відповідальності, план взаємодії учасників, модель звітування за результатами роботи тощо	План взаємодії учасників
2	Регулярні наради	Типологія нарад, моделі проведення нарад, моделі ролей учасників нарад тощо	Усе перелічене
3	Комунікації із зацікавленими сторонами проєкту	Матриця відповідальності, типологія нарад, моделі проведення нарад, моделі ролей учасників нарад тощо	Матриця відповідальності
4	Стратегічні комунікації	Моделі стратегій, матриця відповідальності, типологія нарад, моделі проведення нарад, моделі ролей учасників нарад тощо	Моделі стратегій, матриця відповідальності

Злагоджена робота системи комунікацій у проєкті можлива за умови належного забезпечення керівником проєкту можливості продуктивної взаємодії фахівців у галузі Data Science з управлінцями і фахівцями супутніх галузей проєкту. Важливу роль у цьому також відіграватиме коректне використання моделей колективної ефективності, що застосовують дослідження психології команд сумісно з урахуванням психології кожного їхнього учасника.

Гнучкі інструменти та фреймворки управління проєктом. Сучасний світ IT-проєктів зафондував особливий підхід до їхнього здійснення, заснований на концепті гнучкості і втілений у

методології гнучкого управління Agile. Усе ще залишаючись у тренді проєктного менеджменту, Agile-підхід, тим не менш, яскраво відмінний від нього і розвивається відносно самостійно і впевнено паралельно з основним трендом. Залишаючи за дужками дискусію щодо принципової новизни моделей і методів, що використовуються в Agile, виокремимо ті з них, що застосовні до проєкту створення IT-компанії у галузі Data Science (табл. 6).

Таблиця 6
Моделі та методи гнучкого управління

№	Аналіз моделей/методів		
	Моделі/методи Agile	Різновиди моделей/методів	Застосовність до проєкту
1	Моделі управління проєктом (фреймворки)	Scrum, Kanban, Lean, XP, кастомізовані гібриди (Scrumban etc.) тощо	Кастомізований гібрид Scrumban
2	Моделі життєвого циклу	Інкрементна, ітераційна, V-подібна, комбінована тощо	Інкрементна
3	Моделі ролей у команді проєкту	Відповідно до Scrum, Kanban, XP, 6 сигм тощо	Модель команди за методологією 6 сигм
4	Моделі визначення тривалості робіт	Відповідно до Scrum (скрам покер), Kanban (рамкові тривалості), метод PERT тощо	Модель визначення тривалості за Kanban (рамкові тривалості)
5	Моделі контролю над виконанням робіт	Методи управління якістю, метод Earned Value, календарно-сітьові методи тощо	Календарно-сітьові методи

У спробах у загальному вигляді характеризувати інтегрованість гнучких методів у практику управління IT-проєктами багато дослідників не уникають стереотипу щодо довершеності саме гнучких методів і «впевненої застарілості класичних. Однак у протигагу йому варто зазначити, що масштабні IT-продукти створюються, як правило, з використанням класичних методологій проєктного менеджменту, у межах яких життєвий цикл є водоспадним. Із нез'ясованих причин апологети Agile вважають, що модель життєвого циклу – це єдина модель, що входить до класичних методологій управління, хоча це, звісно, зовсім не так, тому вони помилково асоціюють будь-яку з класичних методологій (яких, до речі, декілька, як їм знов-таки чомусь невідомо) із waterfall методологією. Хоча і той самий PMBOK і PRINCE2, і P2M пропонують широкую множину



інших моделей управління – командою, часом реалізації, вартістю, змістом, цінностями тощо.

Модель вибору множини моделей і методів у межах визначених ключових галузей знань для досліджуваного проекту пропонується у вигляді згортки критеріїв:

$$C = \mu(a_i, b_j, c_k, d_l, e_m, f_n) \times \\ \times \max(a_i) + \max(b_j) + \max(d_l) + \\ + \max(e_m) + \max(f_n) \rightarrow \max,$$

де C – експертна оцінка рейтингу комбінації моделей і методів, що будуть обрані для досліджуваного проекту;

a, b, c, d, e, f – галузі проектного менеджменту, що розглядаються у проекті створення ІТ-компанії у галузі Data Science (а саме галузі управління командою, управління цінностями, розроблення концепції, управління змістом, управління комунікаціями, гнучкі інструменти управління); $a_i, b_j, c_k, d_l, e_m, f_n$ – обраний метод відповідної галузі; $\mu(a_i, b_j, c_k, d_l, e_m, f_n)$ – експертна оцінка рейтингу комбінації обраних методів (оцінка їхнього потенційного синергетичного ефекту); $\max(a_i)$ – пошук моделі або методу з a -ї галузі знань із максимальним рейтингом придатності до досліджуваного проекту, оціненим експертно.

Така модель може бути використана для будь-якого набору галузей знань проекту створення ІТ-компанії у галузі Data Science. Як приклад наведено модель, застосовану до обраного набору галузей, які автори вважають ключовими для забезпечення збільшення ймовірності успіху такого проекту (управління командою, управління цінностями, розроблення концепції, управління змістом, управління комунікаціями, гнучкі інструменти управління).

SWOT-АНАЛІЗ ЗАПРОПОНОВАНОГО У ДОСЛІДЖЕННІ ПІДХОДУ

Підбиваючи підсумки щодо цінності і практичної значущості результатів проведених досліджень, проведемо SWOT-аналіз, виокремивши сильні, слабкі сторони, можливості та загрози, що пов'язані з їхнім використанням (табл. 7).

Таким чином, за результатом SWOT-аналізу, можна сформулювати такі висновки.

Основною сильною стороною запропонованого підходу можна визначити наукову обґрунтованість визначеної множини апробованих методів для досліджуваного проекту.

Серед слабких сторін варто акцентувати увагу на неврахуванні спротиву, який буде чинити впровадженню запропонованого підходу середня

ланка менеджменту. Однак розроблення щодо подолання зазначеного можуть бути розглянуті як перспективні подальші дослідження в обраному напрямку.

Таблиця 7
SWOT-АНАЛІЗ ЗАПРОПОНОВАНОГО ПІДХОДУ

Strengths (сильні сторони)	Weakness (слабкі сторони)
1) Наукова обґрунтованість кожного методу та їхні комбінації для досліджуваного проекту; 2) гнучкість підходу завдяки використанню експертного оцінювання для кожного конкретного проекту; 3) урахування в підході ключових галузей управління проектами.	1) Недостатнє врахування нових напрацювань у галузі управління командою – емпатії, емоційного інтелекту тощо; 2) метод експертного оцінювання має ряд вад, які, у разі некоректного використання методу, завдають отримати адекватний результат; 3) неврахування спротиву середньої ланки менеджменту.
Opportunities (можливості)	Threats (загрози)
1) Створення сучасної, гнучкої, інноваційної компанії на базі науково обґрунтованого підходу; 2) забезпечення масштабування компанії завдяки опису процесів управління проектом створення компанії у галузі Data Science; 3) подальше поглиблення наукових досліджень у визначеному напрямку на базі запропонованих підходів.	1) Несприйняття наукового підходу практиками, що будуть реалізовувати проект; 2) недостатнє підлаштування моделей і методів під конкретний проект; 3) високі темпи змін у галузі можуть призвести до швидкого старіння запропонованого підходу.

Зокрема, у таких дослідженнях може бути розглянута модель реалізації організаційних змін Дж. Коттера [14] у складі восьми кроків: 1) створити в організації загальне розуміння неможливості подальшого управління через застарілі моделі, переконати персонал у необхідності організаційних змін; 2) сформувати команди так званих реформаторів по можливості з найбільш впливових співробітників, усіляко заохочувати діяльність цих команд; 3) створити бачення бажаного майбутнього організації з метою підвищення активності персоналу, розробити стратегію досягнення бачення; 4) активно пропагувати серед персоналу майбутній вигляд організації, використовуючи ефективні методи переконання; 5) створити необхідні умови для втілення організаційних змін у життя – усунути можливі перешкоди, змінити структури й обов'язки, що суперечать новій парадигмі організації, усіляко заохо-



чувати творчий підхід і готовність до ризику; 6) спланувати і досягти короткострокових результатів; 7) закріпити досягнення і розширити перетворення за рахунок створення атмосфери довіри до нових підходів, поширити успішний досвід по всій організації, виконати кадрові зміни і перестановки; 8) інституалізувати проведені організаційні зміни за допомогою формалізації правил поведінки, запровадити відповідну систему заохочень співробітників за досягнення завдань організаційних змін, формувати відповідні проведені змінам системи цінностей організації.

Поміж загроз особливо виокремлюють високі темпи змін у галузі, які можуть призвести до швидкого старіння запропонованого підходу.

Серед можливостей звертає на себе увагу потенціал запропонованого комплексу моделей щодо створення сучасної, гнучкої, інноваційної компанії на базі науково обґрунтованого підходу.

Загалом, результати SWOT-аналізу свідчать про актуальність і практичну цінність запропонованого підходу та його середню ризиковість, що, в основному, лежить у організаційній площині – спроможності керівника проекту до впровадження змін із подоланням викликів і спротивів, що виникнуть.

Як практичні приклади проектів, у межах яких може бути доцільним застосування запропонованого підходу, можна навести такі: проект створення IT-компанії з аналізу BigData за державним замовленням, проект створення IT-компанії з розроблення спеціалізованих нейронних мереж, проект розроблення міжнародної IT-компанії широкої Data Science спеціалізації.

ВИСНОВКИ

Сучасний діджиталізований і такий, що все далі і глибше діджиталізується, світ висуває конкурентні вимоги до IT-середовища і продукує появу у ньому компаній, що уособлювали б сучасні досягнення і тренди. Слід особливо виокремити галузь Data Science як тренд сучасного IT-середовища останнього десятиліття. Зважаємо на те, що відповідна предметна галузь досліджена і досліджується достатньо глибоко й детально, водночас організаційний контекст, що стосується створення таких компаній, висвітлено у наукових напрацюваннях недостатньо. Застосувавши до такого створення проектний підхід, у цій статті ми розглянули моделі й методи, що можуть бути застосовані у комплексі для зазначеного типу проектів – проектів створення IT-компаній у галузі Data Science. Основні галузі управління відповідним проектом (управління командою, управління цінностями, розроблення

концепції, управління змістом, управління комунікаціями, гнучкі інструменти управління), які було виокремлено й охарактеризовано, забезпечують створення наукового підґрунтя до розроблення системи управління для нього.

Модель вибору множини моделей і методів у межах визначених ключових галузей знань для досліджуваного проекту дозволить у прикладних умовах реалізації здійснити обґрунтований підбір моделей і методів, що будуть оціночно адекватними для відповідної системи управління.

Проведений SWOT-аналіз запропонованого підходу довів його актуальність і практичну цінність та обґрунтував середню ризиковість.

Слід також підкреслити практичну значущість запропонованого підходу з погляду економії часу, який проектний менеджер витратить на підбір моделей і методів управління проектом. Наведена в межах запропонованого підходу систематизація, оціночно, дозволить заощадити 10–15 % часу на підбір доцільних моделей, використання яких у досліджуваному проекті (унаслідок їхньої адекватності) у свою чергу дозволить, оціночно, на 2–3 % скоротити час реалізації таких проектів.

Як перспективи дослідження в обраному напрямі слід виокремити розроблення моделей і методів адаптації (або, можливо, біадаптації [15]) запропонованого підходу до умов реалізації деяких типів досліджуваних проектів, зокрема в аспекті подолання організаційного спротиву менеджменту середньої ланки до запровадження організаційно-управлінських інновацій (в умовах реалізації проекту створення нової IT-компанії в межах існуючої групи IT-підприємств).

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

- [1] A Guide to the Project Management Body of Knowledge (PMBOK® Guide) – Sixth Edition [Текст] / USA. – *Project Management Institute*, 2017. – 537 p.
- [2] Managing Successful Projects with PRINCE2 [Текст] / OGC (Office of Government Commerce). – TSO (The Stationery Office), *Printed in the United Kingdom for TSO*, 2009. – 327 p.
- [3] ISO 21500:2012. Guidance on project management [Текст] / *Project Committee ISO/PC 236*, 2012. – 36 p.
- [4] IPMA Individual Competence Baseline (ICB) Version 4.0 for Project, Programme & Portfolio Management [Текст] / *IPMA*. – *Nijkerk, Netherlands*. – 2015. – 431 p.
- [5] Martin, R. C. Agile Software Development: Principles, Patterns, and Practices [Текст] / Robert C. Martin. – *Pearson*. – 2002. – 528 p.
- [6] Agile Practice Guide: Paperback [Текст] / USA. – *Project Management Institute*. – 2017. – 210 p.
- [7] The Standard for Project Management and a Guide to the Project Management Body of Knowledge (PMBOK® Guide) – Seventh Edition [Текст] / USA. – *Project Management Institute*, 2021. – 250 p.
- [8] P2M. Program & Project Management for Enterprise Innovation [Електронний ресурс]. – *Project Management Association of Japan*, 2016. – Режим доступу: http://www.pmaj.or.jp/ENG/p2m/p2m_guide/p2m_guide.html.



[9] Лецер Ю. О. Формування комплексної економіко-організаційної моделі підприємств з розробки інформаційних технологій [Текст]: дис. ... канд. економ. наук : 08.00.04 / Лецер Юрій Олександрович. – Рівне : НУБГП, 2018. – 222 с.

[10] Сокол К. М. Ітераційна модель формування стратегії міжнародного маркетингу [Електронний ресурс] / К. М. Сокол // *Інфраструктура ринку: електрон. наук.-практ. журн.* – 2018. – Вип. 21. – С. 363–368. – Режим доступу: http://www.market-infr.od.ua/journals/2018/21_2018_ukr/61.pdf.

[11] Morozov V., Kalnichenko O., Kolomiets A.: Research of the Impact of Changes Based on External Influences in Complex IT Projects [Текст] // *Proceedings of the IEEE International Conference on Advanced Trends in Information Theory, ATIT 2019*, 18-20 December Kyiv, Ukraine, 2019. – pp. 481–488.

[12] Morozov V., Mezentseva O., Proskurin M. Trainable neural networks modelling for a forecasting of start-up product

development [Текст] // *Proceedings of the 2020 IEEE 3rd International Conference on Data Stream Mining and Processing, DSMP-2020*. – 2020, pp. 55–60.

[13] Неізвесний С. І. Розвиток методологій управління проектами із застосуванням механізмів конвергенції [Текст] : автореф. дис. ... докт. наук з управління проектами та програмами : 05.13.22 / С. І. Неізвесний. – Київ. нац. ун-т будівн. і архітектури, 2013. – 39 с.

[14] Kotter J.P. *Leading Change*. 1st edition [Текст] / J.P. Kotter. – Harvard Business Press, 1996. – 208 p.

[15] Тімінський О.Г. Технології адаптивного управління як механізм забезпечення ефективності організаційно-управлінських систем [Текст] // *Управління розвитком складних систем*. – 2016. – № 27. – С. 122–131.

Стаття надійшла до редколегії

19.10.2021

Project management models for creating an IT company in the field of Data Science

The basis for the emergence of projects to create IT companies in the field of Data Science is considered. The relevance of such projects is substantiated. The feasibility of their study using models and methods of project management is proven. Literary sources in three directions are analyzed. Such areas include classical project management, flexible approaches to project management, value-oriented management. Insufficient research of the described subject is proved. Groups of models and methods are described that describe the relevant areas of project management knowledge that will be necessary and minimally sufficient to develop the scientific basis of the IT company creation project in the field of Data Science. Namely: project team management, project value management, project concept development, project content management, project communications management, flexible project management tools. In the analysis of each area, the relevant models and methods are highlighted. Their applicability to the researched project is analyzed. A model for selecting a set of models and methods within the identified key knowledge areas for the research project in the form of a convolution of criteria is proposed. SWOT-analysis of the proposed approach was performed. The strengths, weaknesses, opportunities and threats associated with the proposed approach are highlighted. Conclusions to the study are formulated. Prospects for further research in the chosen direction are outlined.

Keywords: information technologies, project management, Data Science, IT company.



Олександр Тімінський. Працює доцентом кафедри технологій управління факультету інформаційних технологій Київського національного університету імені Тараса Шевченка, канд. техн. наук, Київ, Україна. Спеціаліст з автоматизованого управління технологічними процесами і виробництвами, Закінчив Київський національний університет будівництва і архітектури 1994 р., де здобув ступінь канд. техн. наук 2001 р. З 2014 р. викладає на факультеті інформаційних технологій Київського національного університету імені Тараса Шевченка. Київ, Україна. Автор понад 100 наукових робіт, зокрема двох посібників і монографії.

Alexander Timinsky. Candidate of technical sciences (PhD), Associate Professor of the Department of Management Technologies, Faculty of Information Technology, Taras Shevchenko National University of Kyiv, Ukraine. He received a degree in automated control of technological processes and production at the Kyiv National University of Construction and Architecture in 1994. He also received the degree of Candidate of Sciences in 2001. Since 2014 he has been teaching at the Faculty of Information Technologies of the Taras Shevchenko National University of Kyiv. Kyiv, Ukraine. Author of more than 100 scientific works, including 2 textbooks and 1 monograph.



Анна Коломієць. Працює доцентом кафедри технологій управління факультету інформаційних технологій Київського національного університету імені Тараса Шевченка, Київ, Україна. Закінчила Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського» 2011 р., магістр з економічної кібернетики. Здобула ступінь канд. екон. наук у Національному технічному університеті України «Київський політехнічний інститут імені Ігоря Сікорського» 2015 р. З 2015 р. викладає на факультеті інформаційних технологій Київського національного університету імені Тараса Шевченка. Київ, Україна. Автор понад 60 наукових робіт, зокрема двох посібників та двох монографій.

Anna Kolomiets. Candidate of Economic Sciences (PhD), Associate Professor of the Department of Management Technologies, Faculty of Information Technologies, Taras Shevchenko National University of Kyiv, Ukraine. She received a master's degree in economic cybernetics from the National Technical University of Ukraine "Kyiv Polytechnic Institute named after Igor Sikorsky" in 2011. She received the degree of Candidate of Sciences at the National Technical University of Ukraine "Kyiv Polytechnic Institute named after Igor Sikorsky" in 2015. Since 2015 he has been teaching at the Faculty of Information Technologies of the Taras Shevchenko National University of Kyiv. Kyiv, Ukraine. Author of more than 60 scientific papers, including 2 textbooks and 2 monographs.



Ольга Мезенцева. Працює доцентом кафедри технологій управління факультету інформаційних технологій Київського національного університету імені Тараса Шевченка, Київ, Україна. Закінчила Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського» 2010 р., магістр з економіки підприємства. Здобула ступінь кандидата наук в Національному технічному університеті України «Київський політехнічний інститут імені Ігоря Сікорського» 2015 р. З 2016 р. викладає на факультеті інформаційних технологій Київського національного університету імені Тараса Шевченка. Автор понад 60 наукових робіт, зокрема посібника та монографії.

Olga Mezentseva. Candidate of Economic Sciences (PhD), Associate Professor of the Department of Management Technologies, Faculty of Information Technology, Taras Shevchenko National University of Kyiv. Kyiv, Ukraine. She received a master's degree in business economics from the National Technical University of Ukraine "Kyiv Polytechnic Institute named after Igor Sikorsky" in 2010. She received the degree of Candidate of Sciences at the National Technical University of Ukraine "Kyiv Polytechnic Institute named after Igor Sikorsky" in 2015. Since 2016 he has been teaching at the Faculty of Information Technologies of the Taras Shevchenko National University of Kyiv. Author of more than 60 scientific papers, including 1 textbook and 1 monograph.

Наукове видання



СУЧАСНІ ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

№ 1 (1) 2021

Друкується за авторською редакцією

Автори опублікованих матеріалів несуть повну відповідальність за підбір, точність наведених фактів, цитат, економіко-статистичних даних, власних імен та інших відомостей. Редколегія залишає за собою право скорочувати та редагувати подані матеріали.

Оригінал-макет виготовлено Видавничо-поліграфічним центром "Київський університет"
Виконавець *В. Гаркуша*



Формат 60x84^{1/16}. Ум. друк. арк. 11,04. Наклад 100. Зам. № 221-10347.
Гарнітура Arial. Папір офсетний. Друк офсетний. Вид. № ІТ6.
Підписано до друку 22.12.2021

Видавець і виготовлювач
ВПЦ "Київський університет",
6-р Тараса Шевченка, 14, м. Київ, 01601, Україна
☎ (38044) 239 32 22; (38044) 239 31 72; тел./факс (38044) 239 31 28
e-mail: vpc_div.chief@univ.net.ua; redaktor@univ.net.ua
<http://vpc.knu.ua>

Свідоцтво суб'єкта видавничої справи ДК № 1103 від 31.10.02