

А. В. Пукас, д-р техн. наук, проф.; А. Ю. Сімак

МЕТОД ФОРМУВАННЯ ПРОГНОЗОВАНИХ ПОКАЗНИКІВ АКТИВНОСТІ НАУКОВО-ПЕДАГОГІЧНИХ ПРАЦІВНИКІВ ДЛЯ ІНТЕЛЕКТУАЛІЗОВАНОЇ СИСТЕМИ АНАЛІЗУ ЇХ ДІЯЛЬНОСТІ ТА РЕЙТИНГУВАННЯ

У сучасному науковому середовищі ефективність дослідників та викладачів часто оцінюється за допомогою формальних показників, таких як кількість публікацій, цитувань, грантів, педагогічних досягнень тощо. У науковій роботі описано метод формування прогнозованих показників активності науково-педагогічних працівників для побудови персоналізованих рекомендацій, що забезпечить підвищення ефективності їх діяльності у майбутньому періоді. Цей метод запропоновано реалізувати як модуль у складі інтелектуалізованої системи аналізу діяльності та рейтингування науково-педагогічних працівників закладу вищої освіти. Прогнозовані показники активності користувача системи можна сформувати на основі аналізу активності інших користувачів за подібною діяльністю. Запропоновано комбінований підхід, який враховує Евклідову відстань, косинусну подібність, спільні та унікальні критерії, бальну відмінність між спільними критеріями, нормалізацію показників та часову динаміку прогресу, а також дозволяє знаходити користувачів зі схожими, але кращими показниками діяльності, і формулювати конкретні пропозиції для покращення результатів діяльності користувача. Взаємодію з даними у роботі реалізовано на основі графової бази даних Neo4j, яка забезпечила ефективне моделювання складних взаємозв'язків між користувачами, їх звітами, критеріями, даними користувачів, бальними показниками та результатами. Графова модель дозволяє легко додавати нові типи даних та зв'язки, що робить процес пошуку масштабованим та гнучким, та аналіз результатів більш простішим. Аналіз часової динаміки (загальний прогрес, щорічний приріст балів) додає контекст до порівняння користувачів, що збільшує точність знаходження схожих користувачів. Обчислення загального індексу подібності через вагову суму різних факторів (спільні критерії, унікальні критерії, покращення балів, Евклідова відстань, косинусна подібність) забезпечує комплексний підхід до аналізу.

Ключові слова: інтелектуалізована система, системний аналіз діяльності науково-педагогічних працівників, рейтингування, метод формування показників, Евклідова відстань, косинусна подібність, часові ряди, нормалізація, індекс подібності, графова база даних.

Вступ

У сучасному науковому середовищі ефективність дослідників та викладачів часто оцінюється за допомогою формальних показників, таких як кількість публікацій, цитувань, грантів, педагогічних досягнень тощо. Ці показники зазвичай збираються у різному форматі, наприклад, це можуть бути річні звіти, які потім аналізуються адміністративним персоналом або спеціалізованими комітетами. Проте такий підхід має значні обмеження: він є трудомістким, може містити дані різного формату, потребує спеціалізованих працівників для роботи з ними, та є складним для аналізу і надання чітких рекомендації для подальшого розвитку користувачів.

У Західноукраїнському національному університеті (ЗУНУ) впроваджена у 2017 р. інтелектуалізована система для аналізу діяльності та рейтингування науково-педагогічних працівників (НПП) [1]. Система містить дані про користувачів, кафедри, факультети, науково-педагогічну діяльність працівників, а також рейтинги та статистику їхньої діяльності. Вона дозволяє створювати звіти, які базуються на наборі критеріїв, що описують і охоплюють різні групи активностей НПП. Критерії, які додаються адміністраторами, можуть мати різні способи структурування, формати даних та методи нарахування балів, які користувачі отримують за їх виконання. У системі реалізовано різні рівні користувацьких ролей, які визначають доступ до даних і функціональних можливостей. Щорічно користувачі заповнюють звіти про свою діяльність і отримують бали за відповідні критерії [2].

Кожен критерій, що додається адміністраторами, супроводжується його описом із зазначенням формату подання даних і бальною оцінкою. Ці параметри визначаються експертною комісією та затверджуються вченою радою університету. Такий формат представлення критеріїв дозволяє описати практично будь-яку діяльність користувача та додати її в річний звіт. Система зберігає всі річні звіти користувачів, проте, аналізує лише дані за останні п'ять років [3].

Процедура заповнення звіту є автоматизованою: вона починається з перевірки на валідність введених даних (залежно від типу та формату критерію) і завершується автоматичним збором та формуванням певних даних користувачів. У системі передбачено процес перевірки та рецензування звітів завідувачами кафедр та рецензентами. Після перевірки та затвердження звіту користувач отримує бали за річний звіт, які враховуються у загальних рейтингах [4]. Зображення затвердженого звіту користувача з рейтингової системи університету, зображено на рис. 1.

Звіт Пукас Андрій Васильович за 2024 рік

Звіт затверджено!

Загальний бал: 1800.00

*** Досягнута кваліфікація ***

Всього балів: 50

1. Наявність наукового ступеня [макс. бал=30]

☒ Доктор наук [30]

☐ Кандидат наук / Доктор філософії [20]

Бал = 30.00/30

2. Наявність вченого звання [макс. бал=20]

☒ Професора [20]

☐ Доцента [10]

☐ Старшого наукового співробітника [10]

☐ Провідного наукового співробітника [20]

Бал = 20.00/20

*** Публікаційна активність ***

Всього балів: 857.14

16. Публікації у співавторстві у міжнародних наукових виданнях, які входять до наукометричних баз Scopus чи Web of Science без квартилю [N°200/K]

1) Pukas Andriy, Simak Andriy, Yushko Andriy, Nadvynychnyy Sergiy, Shandruk Serhii, Vatslavskiy Oleh, "Automated Reporting Module in the Academic Staff Performance Appraisal System" IEEE, 2024

K = 6

<http://dx.doi.org/10.1109/acit62333.2024.10712573>

1) Sysioleva Inna, Pukas Andriy, Pohrishchuk Oleh, Pohrishchuk Borys, Tsikhanovska Olena, Lyzun Maria, "Artificial Intelligence as an Organized Assembly of Information Technologies for the Goals of Sustainable Development" IEEE, 2024

Рис. 1. Сторінка перегляду затвердженого звіту користувача

Оцінка за науково-педагогічну активність користувачів варіюється: деякі мають високі рейтинги, інші – низькі. Підвищення ефективності діяльності НПП сприяє зростанню їхніх рейтингів у системі, що, у свою чергу, позитивно впливає на загальний рейтинг закладу вищої освіти. Для підвищення ефективності діяльності працівників необхідно розробити метод формування рекомендацій, який допоможе покращити їхню активність. Рекомендації повинні базуватись на основі аналізу і порівняння діяльності інших НПП, та враховувати подібні умови праці, групи активностей, кваліфікацію та інше.

Мета і завдання статті

Метою роботи є підвищення ефективності діяльності НПП на основі розробки методу формування прогнозованих показників їх активності для інтелектуалізованої системи аналізу діяльності та рейтингування. Вказаний метод дозволить:

- аналізувати діяльність користувачів у різних групах активностей (публікації, гранти, педагогічна діяльність тощо);
- визначати користувачів зі схожими сценаріями діяльності, але кращими показниками;
- формувати персоналізовані пропозиції для покращення науково-педагогічної діяльності на основі аналізу кращих практик.

Завдання дослідження:

1. Проаналізувати відомі методи формування персоналізованих рекомендацій, визначити їхні переваги та недоліки, дослідити складність реалізації та оперування, гнучкість модифікації та обсяг необхідних даних.
2. Розробити метод та алгоритм для знаходження користувачів зі схожими сценаріями діяльності, але кращими показниками на основі використання принципів і методів машинного навчання в широкому сенсі, адаптувавши їх під поточну задачу.
3. Створити ефективну модель для зберігання та представлення даних про користувачів та їх діяльність. Модель повинна враховувати різні аспекти наукової діяльності, такі як: публікаційна діяльність (кількість публікацій, цитувань, індекс Хірша тощо); грантова діяльність (кількість проєктів, обсяги фінансування тощо); педагогічна діяльність (кількість курсів, студентів, наукових керівництв тощо); організаційна діяльність (участь у конференціях, рецензування, наукове керівництво тощо). Модель повинна забезпечувати швидкий доступ до даних, їх демонстрацію та можливість аналізувати складні взаємозв'язки між користувачами.
4. Провести оцінку ефективності запропонованого підходу на реальних даних. Наприклад, це може бути зроблено за допомогою експертної та метричної оцінки точності та повноти знайдених схожих користувачів.

Аналіз літератури

Останніми роками активно розвиваються інтелектуалізовані системи, які можуть аналізувати великий обсяг даних та пропонувати шляхи покращення результатів [5, 6]. Проте більшість наявних систем оцінки та ранжування надають лише загальні показники, без конкретних рекомендацій для подальшого розвитку [7]. Зокрема, загальні показники (наприклад, індекс Хірша) дають лише оцінку активності НПП, але не показують, як саме можна покращити цей показник; річні звіти часто аналізуються лише адміністративним персоналом, без надання користувачам детальних порад для покращення результатів, не враховуючи багатогранності наукової діяльності, яка охоплює публікаційну, грантову, педагогічну, організаційну та інші активності, які можуть враховуватись для підвищення рейтингу як працівника, так й університету.

У сучасній науковій практиці зазвичай використовуються певні технологічні підходи для аналізу та покращення продуктивності дослідників. Розглянемо ключові технологічні рішення, які застосовуються в наукових установах.

Класичні системи керування реляційними базами даних (RDBMS) залишаються основою для зберігання структурованої інформації про науковців. Система рейтингування ЗУНУ реалізована на основі системи керування реляційними базами даних – MySQL [2]. Системи на базі Microsoft SQL Server чи Oracle Database широко використовуються для документування наукових даних, грантів та метрик ефективності. Проте цей підхід має обмеження у виявленні складних взаємозв'язків між різними аспектами наукової діяльності.

Сучасні методики машинного навчання, такі як алгоритми k-NN, кластеризація та нейронні мережі активно впроваджуються в наукових платформах (Google Scholar, ResearchGate та інші) [8]. Ці технології ефективно працюють для персоналізованих рекомендацій літератури чи потенційних співробітників, але часто потребують значних об'ємів тренувальних даних і не охоплюють всі компоненти наукової діяльності.

Спеціалізовані наукометричні платформи або сервіси (Scopus, Web of Science та інші) надають поглиблені можливості для аналізу наукового впливу через системи цитувань.

Однак їх функціонал обмежується переважно публікаційною активністю, не враховуючи інші важливі аспекти професійної діяльності [9].

Графові бази даних, наприклад Neo4j чи ArangoDB, пропонують інноваційний підхід до моделювання складних взаємозв'язків між науковцями та їх досягненнями. Такі системи забезпечують швидкий пошук аналогів та детальний аналіз зв'язків між різними показниками діяльності, що дає можливість створення ефективних рекомендаційних механізмів [10 – 11].

При всій ефективності наявних рішень, багато з них не враховують комплексного характеру наукової діяльності, де, окрім загальних показників, важливим є детальний аналіз окремих напрямів роботи – від науково-дослідної до педагогічної.

Опис переваг та недоліків різних підходів для знаходження схожих користувачів [12 – 13], а також додаткові фактори, такі як обсяг даних, складність реалізації та гнучкість у внесенні змін наведено у таблиці 1.

Таблиця 1

Опис переваг і недоліків різних підходів для знаходження схожих користувачів

Підхід	Переваги	Недоліки	Обсяг даних	Складність	Гнучкість
Класичний k-NN	Простий у реалізації. Широко використовується. (Можна використати як доповнюючий метод)	Чутливий до масштабу ознак. Не враховує часову динаміку. Важко обробляти розрізнені дані (розрізнені групи/критерії)	Середній	Середня	Висока
Регресійні моделі	Здатні прогнозувати майбутні результати. Враховують часову динаміку.	Вимагають велику кількість тренувальних та точних даних. Складне налаштування параметрів.	Дуже великий	Висока	Середня (перетренування)
Статичні порівняння	Прості у реалізації. Швидкі обчислення.	Не враховує часову динаміку. Неможливо виявити прогрес або регрес.	Малий	Низька	Низька
Графовий підхід з часовою динамікою	Враховує часову динаміку та взаємозв'язки між користувачами. Здатний аналізувати складні структури даних.	Вимагає спеціалізованого програмного забезпечення (наприклад, Neo4j). Великі графи – вимагають об'ємних запитів.	Середній	Середня	Висока
Глибинне навчання	Здатні виявляти складні закономірності у даних. Висока точність при достатньому обсягу точних даних.	Вимагає великого обсягу тренувальних даних. Складність налаштування моделей. Вимагає значних обчислювальних ресурсів, перенавчання	Великий	Висока	Висока
Кластеризація	Здатна групувати користувачів за схожими характеристиками. Використовується для широкого класу задач. (Можна використати як доповнюючий метод)	Вимагає попереднього визначення кількості кластерів. Чутлива до якості даних.	Середній	Середня	Середня
Дерево рішень	Легко інтерпретувати результати. Добре працюють для задач класифікації. Можуть обробляти категоріальні та числові дані.	Надмірна залежність від тренувальних даних. Малі зміни в даних можуть призвести до значних змін у структурі дерева.	Великий	Висока	Середня
Випадковий ліс	Ансамблевий підхід знижує ризик перенавчання. Менш чутливий до шуму та аномалій в даних. Ефективно обробляє великі набори даних.	Складність інтерпретації. Вимагає багато часу для навчання на великих наборах даних. Гірше працює з розрідженими даними. Вимагає чітких правил і багатьох прикладів	Великий	Висока	Середня

Продовження таблиці 1					
Метод опорних векторів	Добре працює з великими кількостями ознак. Може обробляти шум та аномалії в даних. Менше підлягає перенавчанню за рахунок регуляризації.	Вимагає правильного налаштування параметрів (наприклад, ядро, параметр C). Вимагає багато часу для навчання на великих наборах даних. При змінах, вимагає перенавчання	Великий	Висока	Середня

Проаналізувавши наявні підходи, та врахувавши особливості наявної рейтингової системи університету, запропоновано комбінований метод знаходження схожих за діяльністю користувачів на основі графового підходу з часовою динамікою та методів машинного навчання у широкому сенсі. Запропонований метод враховує часову динаміку та взаємозв'язки між користувачами, не вимагає великих обсягів даних для початкового аналізу, обчислює конкретні метрики, є гнучким та масштабованим.

Метод та алгоритм знаходження схожих за діяльністю користувачів

Знаходження схожих користувачів є ключовим етапом у побудові рекомендаційного модуля в інтелектуалізованій системі аналізу діяльності та рейтингування НПП. Для цього використано комплексний підхід, який включає принципи і методи машинного навчання, які адаптовані для конкретної задачі, щоб враховувати різні аспекти діяльності користувачів та їх взаємозв'язки. Крім того, для збільшення точності враховується не тільки поточний стан користувача, але й часові ряди. Це включає аналіз прогресу протягом декількох років через обчислення загального прогресу та середнього щорічного покращення [14].

Метод побудови запитів для знаходження кращих сценаріїв діяльності схожих користувачів включає 8 етапів, які наведено нижче.

1. Обчислення метрик активності користувача. Кожен користувач має певну кількість щорічних звітів, які включають дані про його наукові досягнення, розділені на різні групи активностей (публікації, гранти, педагогічна діяльність тощо), тому можна визначити прогрес користувача за роками. Сума балів за всі критерії в звіті формує загальний бал користувача за певний рік. Для аналізу прогресу користувача за визначений звітний період використовуються наступні метрики.

а) Вектор загального рейтингу користувача, як часовий ряд балів користувача за всі роки:

$$\overrightarrow{Stotal^{(u)}} = \{S_1^{(u)}, S_2^{(u)}, \dots, S_T^{(u)}\}, \quad (1)$$

де: $S_t^{(u)}$ – це загальний бал користувача u за рік t , T – рік найстарішого звіту, як правило не більше 5.

б) Вектор балів по групах активностей користувача, як часовий ряд балів користувача за всі роки по кожній групі:

$$\overrightarrow{Sga^{(u)}} = \{s_1^{(u,t_1)}, s_2^{(u,t_1)}, \dots, s_g^{(u,t_1)}, \dots, s_1^{(u,T)}, \dots, s_g^{(u,T)}\} \quad (2)$$

де: $s_g^{(u,T)}$ – бал користувача u за групу активності g за рік t , T – рік найстарішого звіту, g – кількість груп активностей. Сьогодні в наявній системі рейтингування діяльності НПП ЗУНУ є шість груп активностей.

в) Темп динаміки активності користувача – усереднене значення щорічної динаміки балів, яке демонструє відносний темп динаміки активності, який розраховується за умови, що користувач має мінімум 2 заповнених річних звіти:

$$Y^{(u)} = \frac{1}{T-1} \sum_{t=2}^T \frac{(S_t^{(u)} - S_{t-1}^{(u)})}{S_{t-1}^{(u)} + e}, \quad (3)$$

де: $S_t^{(u)}$ – загальний бал користувача u за рік t , T – кількість років, за які користувач u має

звіти, e – мале число (наприклад, 0.01), щоб уникнути ділення на 0. Якщо користувач має тільки 1 заповнений річний звіт, то Y не буде обчислюватись. Низьке значення цього показника вказує на низьку динаміку. Ця метрика дозволяє оцінити наскільки швидко користувач покращує свої результати протягом визначеного часу. Тоді як машинне навчання використовує моделювання трендів через регресійні або інші складні моделі [14, 15].

г) Показник наближеності двох користувачів за темпом динаміки активності:

$$Y_{(u_1, u_2)} = \frac{|Y^{(u_1)} - Y^{(u_2)}|}{\max(Y^{(u_1)}, Y^{(u_2)}) + e} \quad (4)$$

де: $Y^{(u_1)}, Y^{(u_2)}$ – темп зростання за рік користувача u_1, u_2 , e – мале число для уникнення ділення на 0.

2. Обчислення Евклідової відстані для порівняння користувачів. Ця відстань використовується для вимірювання подібності між двома користувачами на основі їх показників у різних групах активностей (публікації, гранти, педагогічна діяльність тощо). Показники користувача представляється у вигляді вектора, де кожна компонента відповідає значенню показника у певній групі активності за певний звітний рік [16].

Формула для обчислення Евклідової відстані між користувачами по групах активності:

$$EvkG_{(u_1, u_2)} = \sqrt{\sum_{t=1}^T \sum_{g=1}^n \left(s_g^{(u_1, t)} - s_g^{(u_2, t)} \right)^2}, \quad (5)$$

де: $s_g^{(u_1, t)}$ – бал користувача u_1 за групу активності g за рік t , $s_g^{(u_2, t)}$ – бал користувача u_2 за групу активності g за рік t , n – кількість груп активностей, g – конкретна група активності, T – всі доступні звітні роки.

Також обчислюємо Евклідову відстань між користувачами по загальному балу:

$$EvkS_{(u_1, u_2)} = \sqrt{\sum_{t=1}^T \left(S_t^{(u_1)} - S_t^{(u_2)} \right)^2}, \quad (6)$$

де: $S_t^{(u_1)}$ – показник загального балу користувача u_1 за рік t , $S_t^{(u_2)}$ – показник загального балу користувача u_2 за рік t , T – всі доступні звітні роки.

Низька Евклідова відстань означає, що користувачі мають схожі показники у багатьох групах, а велика відстань свідчить про значні відмінності у діяльності користувачів.

3. Обчислення конусної подібності користувачів [15, 16]. У машинному навчанні для знаходження подібних сценаріїв активності користувачів часто використовуються складні функції, наприклад, косинусної подібності (Cosine Similarity). Ця подібність вимірює кут між двома векторами у багатовимірному просторі. Оскільки вона корисна для порівняння напрямків, а не абсолютних величин, то її можна використовувати як доповнення до Евклідової відстані.

Формула для обчислення косинусної подібності:

$$CosG_{(u_1, u_2)} = \frac{(\overrightarrow{Sga^{(u_1)}} \cdot \overrightarrow{Sga^{(u_2)}})}{(\|\overrightarrow{Sga^{(u_1)}}\| \cdot \|\overrightarrow{Sga^{(u_2)}}\|)}, \quad (7)$$

де: $\overrightarrow{Sga^{(u_1)}}, \overrightarrow{Sga^{(u_2)}}$ – вектори балів по групах активностей користувачів u_1 та u_2 за всі роки.

Ця подібність краще підходить для аналізу співвідношень між показниками, а саме: як розподілені бали між групами активностей у користувачів, тобто, яку переважну діяльність ведуть користувачі. Чим вищий цей показник, тим більш схожу діяльність ведуть користувачі, що порівнюються. Порівняння косинусної подібності по окремих активностях є недоцільним, тому що: косинусна подібність втрачає точність за високої розмірності; збільшуються шумові критерії (в цілому, менш важливі критерії можуть більше впливати на косинусну подібність); різні масштаби балів вимагають додаткової нормалізації балів кожного з критеріїв; різна кількість активностей у користувачів (збільшує розмірність і вимагає встановлення нулів в відсутні критерії).

4. Обчислення індексу Жаккара (Jaccard Index) [13]. Цей етап здійснюється для зменшення впливу великої кількості унікальних критеріїв одного користувача та отримання нормалізованих показників подібності без додаткових обчислень. Для кожного користувача формується список критеріїв, за якими вони отримали певні бали. Далі проводиться аналіз критеріїв, які є спільними для двох користувачів, а які є унікальними у звітному періоді.

$$J(u_1, u_2) = \frac{\sum_{c \in C(u_1) \cap C(u_2)} \min(s_c^{(u_1)}, s_c^{(u_2)})}{\sum_{c \in C(u_1) \cup C(u_2)} \max(s_c^{(u_1)}, s_c^{(u_2)})}, \quad (8)$$

де: $J(u_1, u_2)$ – індекс Жаккара; $s_c^{(u_1)}, s_c^{(u_2)}$ – бали користувачів u_1 та u_2 за критерій c ; $C(u_1) \cap C(u_2)$, $C(u_1) \cup C(u_2)$ – множини спільних та унікальних критеріїв користувачів u_1 та u_2 .

Чим ближче значення індексу до 1, тим більше подібні активності користувачів.

5. Оцінювання різниці (Difference Score). Для порівняння ефективності двох користувачів за однаковими критеріями використовується метрика різниці балів, обчислена на множині спільних критеріїв. Ця метрика дозволяє виявити, наскільки один користувач перевершує іншого у конкретних аспектах діяльності.

Формула для обчислення сумарної різниці балів за спільними критеріями:

$$\Delta P = \sum_{c \in C(u_1) \cap C(u_2)} (s_c^{(u_2)} - s_c^{(u_1)}), \quad (9)$$

де: $C(u_1, u_2)$ – множина спільних критеріїв, $s_c^{(u_1)}, s_c^{(u_2)}$ – бали користувача u_1, u_2 за критерій c . Цей показник демонструє сумарну перевагу активностей одного користувача над іншим у контексті спільних ознак. Для використання в комплексному індексі подібності цей показник нормалізується. Такий підхід відповідає концепції Difference Score та використовується в аналізі ефектів у задачах машинного навчання, особливо в рекомендаційних системах.

6. Обчислення фактору переваги. Для врахування загальної ефективності користувача вводиться фактор переваги, який базується на порівнянні середньорічних балів двох користувачів, можна розраховувати так:

$$F(u_1, u_2) = \min\left(1, \frac{\overline{s^{(u_2)}}}{\overline{s^{(u_1)}}}\right), \quad (10)$$

де: $\overline{s^{(u_1)}}, \overline{s^{(u_2)}}$ – середньорічний бал користувачів u_1 та u_2 відповідно. Якщо $\overline{s^{(u_2)}} < \overline{s^{(u_1)}}$, то $F(u_1, u_2) < 1$, і користувач u_2 буде в меншому пріоритеті для рекомендації користувачу u_1 .

7. Нормалізація показників. Оскільки різні показники можуть мати різні розмірності, то необхідно нормалізувати всі показники до одного діапазону (0, 100), щоб отримати однаковий формат даних [13].

Загальні формули для нормалізації значень для загального індексу подібності:

$$a) N(x) = \frac{1}{1+|x|}, \quad (11)$$

$$б) N(x) = \frac{x - x_{min}}{x_{max} - x_{min}} * 100, \quad (12)$$

де: $N(x)$ – нормалізоване значення показника, x – абсолютне значення показника (бал, різниця, подібність тощо), x_{max}, x_{min} – максимальне і мінімальне значення на всій множині.

8. Загальний індекс подібності. Обчислюється як ваговий сумарний загальний показник, який враховує усі вище описані етапи.

Формула для обчислення індексу подібності:

$$\begin{aligned} Sim_{u_1 u_2} = & F(u_1, u_2) \cdot (\lambda_1 \cdot N(EvkS_{(u_1, u_2)}) + \lambda_2 \cdot N(CosG_{(u_1, u_2)}) + \lambda_3 \cdot N(EvkG_{(u_1, u_2)}) + \\ & + \lambda_4 \cdot N(J(u_1, u_2)) + \lambda_5 \cdot N(Y_{(u_1, u_2)}) + \lambda_6 \cdot N(\Delta P)), \end{aligned} \quad (13)$$

де: $\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5 + \lambda_6 = 1$ – вагові коефіцієнти, які можна розділити рівномірно, і прирівняти до 0,2, крім показника наближеності за темпом динаміки активності та показника сумарної переваги, вага яких у загальній подібності є слабшою порівняно з іншими),

$F(u_1, u_2)$ – фактор переваги між користувачами, $N(EvkS_{(u_1, u_2)})$ – нормалізована Евклідова відстань по загальному балу, $N(CosG_{(u_1, u_2)})$ – нормалізована косинусна подібність по балах у групах активності, $N(EvkG_{(u_1, u_2)})$ – нормалізована Евклідова відстань по балах у групах активності, $N(J(u_1, u_2))$ – нормалізований Jaccard Index подібності активності, $N(Y_{(u_1, u_2)})$ – нормалізований показник наближеності за темпом динаміки активності, $N(\Delta P)$ – нормалізована оцінка різниці показників.

Загальний індекс подібності дозволяє проводити ранжування користувачів за їхньою подібністю до поточного користувача, з урахуванням його історичного прогресу по відношенню до інших, але з кращими показниками діяльності, та використовувати їх стратегії розвитку як приклад для рекомендації поточному користувачу. Чим вище значення індексу, тим вищий ранг рекомендованого користувача.

Для випадків рекомендації користувачу мінімальної кількості вперше виконуваних активностей проводимо семантичний аналіз подібності нових критеріїв для користувача. Враховуючи групи активності користувача з найбільшими балами, визначаємо векторне представлення (embedding) опису критеріїв усіх користувачів. Для цього використано модель Sentence-BERT (distiluse-base-multilingual-cased-v2), яка є мультимовною, достатньо швидкою і демонструє високу якість, враховуючи невеликий її розмір:

$$\vec{v}_c = BERT(d_c), \quad (14)$$

де: c – це критерій (наприклад: [id:1; «Вчене звання»]), d_c – текстовий опис за критерієм («Вчене звання»), $\vec{v}_c$ – вектор (embedding) текстового опису за критерієм, отриманий з моделі BERT, $BERT(d_c)$ – функція, яка приймає текстовий опис і повертає його числове представлення у високорозмірному просторі. Модель повертає вектор фіксованої довжини (зазвичай 512–768 елементів), цей вектор кодує семантичне значення опису критерія так, що схожі за змістом критерії мають близькі вектори.

Семантичну подібність між описами критеріїв отримуємо як косинусну подібність:

$$Sim_{sem}(c_i, c_j) = \cos(\vec{v}_{c_i}, \vec{v}_{c_j}), \quad (15)$$

де: $\vec{v}_{c_i}, \vec{v}_{c_j}$ – це вектори критеріїв c_i та c_j , отримані за допомогою BERT, $\cos()$ – косинусна подібність у діапазоні від -1 до 1, відповідно змінюється від протилежних до схожих.

Показник семантичної подібності користувачів за критерієм c , який не виконувався користувачем u_1 , але виконувався користувачем u_2 можна розрахувати так:

$$RecScore(c, u_1, u_2) = \lambda_{c1} \cdot s_c^{(u_2)} + \lambda_{c2} \cdot \max_{c' \in Str(u_1)} Sim_{sem}(c, c') \quad (16)$$

де: $s_c^{(u_2)}$ – бал, який отримав користувач u_2 за критерієм c , $Str(u_1)$ – множина критеріїв користувача u_1 з трьох груп активностей з найвищими балами, $Sim_{sem}(c, c')$ – семантична подібність між критеріями c та c' , λ_{c1} та λ_{c2} – вагові коефіцієнти (семантична подібність опису критеріїв важливіша ніж бал, тому $\lambda_{c1} = 0,2, \lambda_{c2} = 0,8$).

Розроблений метод з використанням мультимовної моделі забезпечує пошук критеріїв, які відповідають сфері діяльності користувача та можуть використовуватись для рекомендацій.

Для підвищення ефективності роботи наведеного вище методу в наявній системі рейтингування діяльності науково-педагогічних працівників університету та зниження складності маніпуляції з даними розроблено нову графову модель даних. Алгоритм для знаходження схожих користувачів, зображений на рис. 2.

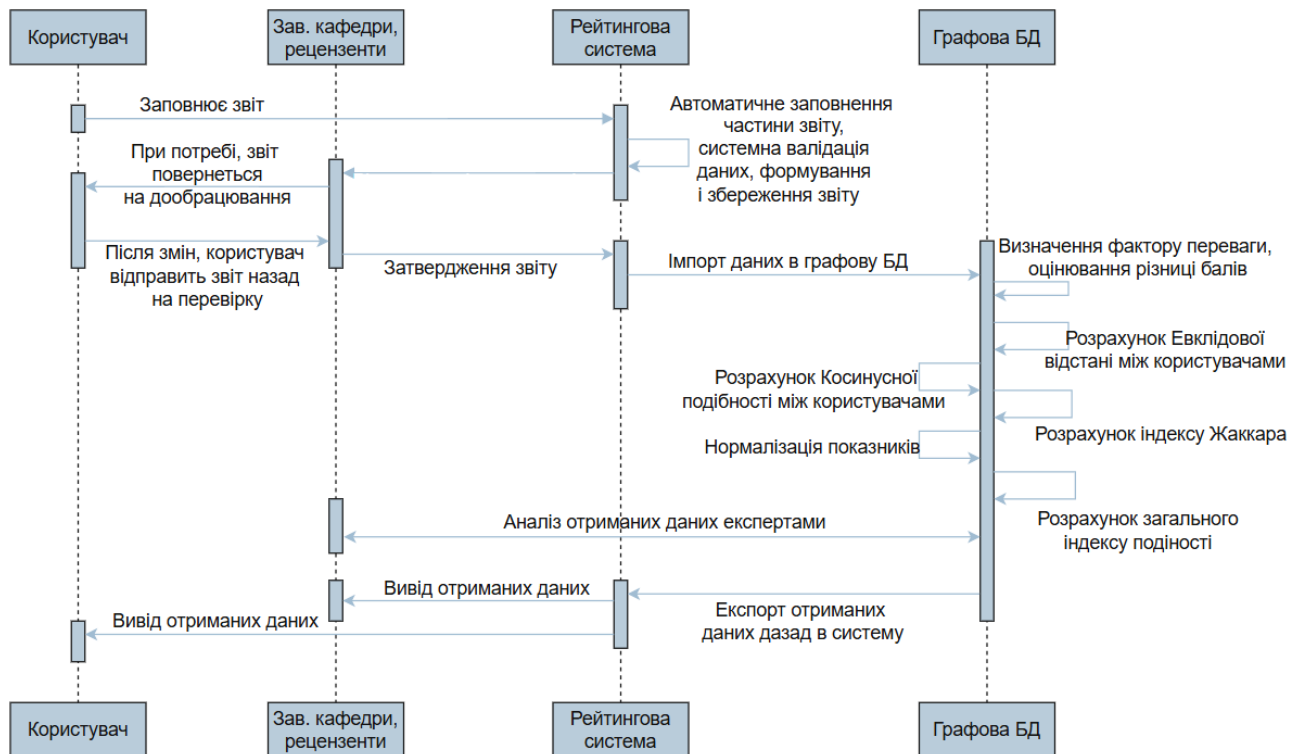


Рис. 2. Діаграма послідовності алгоритму знаходження схожих користувачів

Запропонований підхід має рядом значимих переваг, які роблять його ефективним і практичним для реалізації:

- врахування динаміки розвитку, а саме вміння враховувати не тільки поточний рівень досягнень, але й тренди їх зміни, надає можливість виявляти користувачів з перспективним розвитком;

- мінімальна складність реалізації – графова модель у комбінації з простими математичними методами (наприклад, подібність за косинусом, Евклідової відстані, тощо) не потребує використання складних алгоритмічних структур, таких як нейронні мережі чи регресійні моделі [10];

- висока адаптивність, завдяки архітектурі моделі, яка дозволяє легко інтегрувати нові параметри оцінки (нові категорії показників або типи даних). Наприклад, за допомогою текстової аналітики (TF-IDF, BERT) можливо враховувати тематичну подібність активності користувачів, що розширює функціональні можливості моделі без радикальної перебудови її логіки;

- ефективна обробка запитів у порівнянні з традиційними реляційними базами даних – графова база даних забезпечує значно швидший доступ до пов'язаних даних завдяки прямому представленню зв'язків у вигляді ребер між вузлами, що значно скорочує час виконання запитів порівняно з операціями об'єднання таблиць;

- толерантність до недостатньої повноти даних – навіть при наявності обмежених наборів даних для окремих користувачів модель здатна проводити аналіз на основі доступної інформації та знаходити схожі сценарії діяльності, що зменшує вплив пропущених даних на точність результатів;

- баланс між простотою та функціональністю – застосування адаптованих методів машинного навчання дозволяє отримувати значимі результати без зайвої складності обчислень та практичності реалізації, дозволяючи легко оптимізувати та розширити функціонал системи.

Проектування графової моделі для представлення й аналізу даних

Враховуючи факт, що основні дані діяльності користувачів зберігаються в табличній БД, а Наукові праці ВНТУ, 2025, № 2

формат збереження і представлення даних є складним для аналізу через специфіку наявної інтелектуалізованої рейтингової системи, то для реалізації було використано графову систему керування базами даних Neo4j [11]. Серед її переваг варто називати такі: ефективний аналіз взаємозв'язків між користувачами, їх діяльністю та результатами; швидкий пошук користувачів зі схожими показниками; легке додавання нових типів даних та зв'язків.

На основі поточної бази даних наявної системи рейтингування, для графової моделі було створено модель даних предметної області із наступними вершинами та зв'язками, рис. 3.

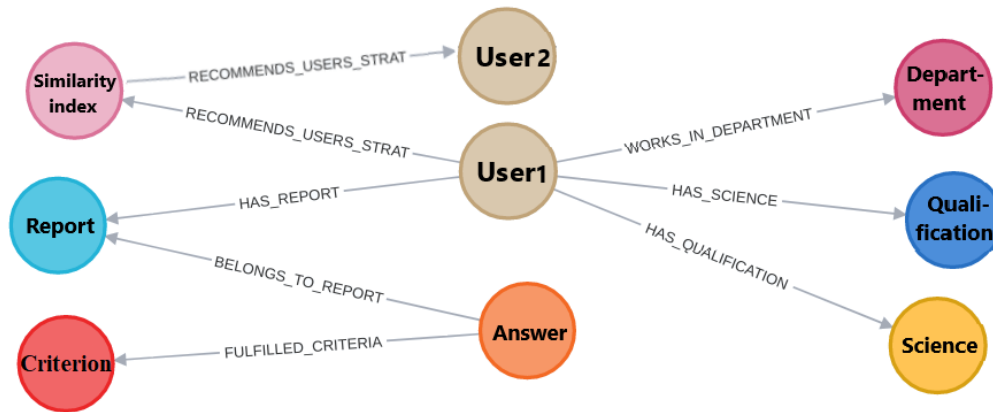


Рис. 3. Графове представлення предметної області

Після імпорту необхідних даних в графову БД можна візуально взаємодіяти та аналізувати діяльність конкретних користувачів у графовому представленні. Приклад діяльності користувача з ID = 581 наведено на рис. 4.

Вершини:

- User: користувач (світло-коричневий елемент). Атрибути: id, name, qualification, science.
- Report: річний звіт користувача (голубий). Атрибути: id, user_id, template_id, totalPoints, year, group activity 1-6.
- Criteria: критерій оцінки діяльності (червоний). Атрибути: id, text, notes, max_point, grAct_id, type.
- Answer: дані користувача за критерієм (оранжевий). Атрибути: id, points, text, options, N, K, report_id, criteria_id.
- Department: кафедра, до якої належить користувач (вишневий).
- Qualification, Science: Кваліфікація та науковий ступінь користувача (жовтий та синій).
- RecommendedUsersOfRecommendationBuilding: показник подібності користувача 1 з користувачем 2 (світло-рожевий). Атрибути: id, year, userId, recommendUserId, indexSimilarity.

Зв'язки:

- HAS_REPORT (User → Report): зв'язує користувача з його звітом.

- PART_OF_REPORT (Criteria $\rightarrow$ Report): зв'язує критерій з шаблоном звіту.
- WORKS_IN_DEPARTMENT (User $\rightarrow$ Department): зв'язує користувача з кафедрою.
- HAS_QUALIFICATION (User $\rightarrow$ Qualification): зв'язує користувача з його кваліфікацією.
- HAS_SCIENCE (User $\rightarrow$ Science): зв'язує користувача з його науковим ступенем.
- FULFILLED_CRITERION (Answer $\rightarrow$ Criteria): зв'язує відповідь з критерієм.
- RECOMMENDS_USERS_STRAT (User $\rightarrow$ RecommendedUsersOfRecommendationBuilding): зв'язує поточного користувача з подібними користувачами.

Визначення ефективності запропонованого підходу є ключовим етапом дослідження, оскільки дозволяє оцінити якість виявлення схожих профілів та практичність рекомендаційних результатів. Для перевірки точності отриманих результатів можна застосувати методичний підхід, базовий на експертній оцінці. Експертам пропонувалось провести аналіз подібності автоматично знайдених користувачів та їх характеристик діяльності, враховуючи доступні метрики та показники. Також передбачається візуальна інтерпретація результатів через графічне представлення взаємозв'язків між користувачами та аналіз структурованих звітів.

11

Для об'єктивної оцінки результатів застосовано метричний підхід, що дозволяє точно вимірювати якість отриманих даних. Приклад, запит для визначення схожості користувачів (відносно особи з ID = 581, яка була розглянута раніше) та отриманий результат представлено на рис. 6.

neo4j\$ MATCH (rsbu:RecommendStrategyBestUsers) where rsbu.userId = 581 01

U2	F_U1U2	EvkG_U1U2	EvkS_U1U2	CosG_U1U2	J_U1U2	Y_U1U2	P	Sim_U1U2
2	50.25	91.04	81.76	94.0	57.69	55.75	52.5	67.54
580	28.29	82.68	83.44	93.06	70.0	56.89	37.5	65.24
789	26.21	81.94	78.71	92.12	51.72	53.66	30.0	57.96
855	30.21	66.55	65.01	94.0	48.39	44.33	25.0	53.21
300	47.68	54.46	61.83	92.12	45.16	42.16	27.5	52.55
563	40.27	52.79	55.98	89.3	51.61	39.76	20.0	51.31
61	23.42	62.12	78.0	89.3	45.16	55.23	15.0	51.15
720	35.1	61.9	48.39	86.48	33.08	26.17	52.5	46.89
57	25.58	45.11	65.67	81.78	45.16	44.77	15.0	45.39
117	15.97	74.17	41.05	82.24	48.48	27.99	0.0	41.69

MAX COLUMN WIDTH:

Рис. 6. Представлення результату знайдених найбільш подібних користувачів з характеристиками, які враховуються в загальному індексі подібності

Результати роботи алгоритму по знаходженню найбільш подібних за діяльністю користувачів відсортовані за індексом подібності, щоб проілюструвати значення різних характеристик, що в кінцевому випадку впливають на цей показник. Користувачі з найвищими характеристиками подібності (виділені зеленим кольором) знаходяться найвище у таблиці. З такого набору отриманих даних використано три найбільш схожі користувачі, а саме їх профілі, для подальшого використання в модулі рекомендацій інтелектуалізованої рейтингової системи аналізу діяльності та рейтингування викладачів, що дозволяє створити персоналізовані рекомендації для кожного користувача.

Після проведеної роботи та аналізу отриманих результатів імпортуємо дані з графової бази в базу даних рейтингової системи. Таким чином відокремлено процес знаходження схожих користувачів від основної системи, розвантажено основну систему від надлишкової роботи, та отримано результуючі, готові до подальшого використання, дані.

Висновки

У цій науковій роботі запропонований метод формування прогнозованих показників діяльності НПП для інтелектуалізованої системи аналізу та рейтингування їх діяльності. Формування таких рекомендацій допомагає покращити результативність діяльності НПП. Запропонований підхід використовує адаптовані методи машинного навчання для аналізу та підбору схожих користувачів по їх діяльності, з урахуванням часових рядів, динаміки активності, семантики критеріїв та фактору переваги, а також графову базу даних для моделювання складних взаємозв'язків між користувачами та їх досягненнями.

Отримані результати показують, що запропонований підхід здатний ефективно знаходити користувачів зі схожими сценаріями діяльності, але кращими показниками по відношенню до поточного користувача. Підхід демонструє ефективні результати у підборі схожих сценаріїв діяльності, які в майбутньому планується використовувати для побудови текстових рекомендацій для користувачів. Використовуючи отримані дані, можна сформулювати такі майбутні напрями досліджень для реалізації модуля рекомендацій в інтелектуалізованій системі аналізу діяльності та рейтингування викладачів:

- розширення модуля для врахування додаткових факторів, таких як спеціалізація користувача, наукові інтереси, врахування контекстуальної інформації, тематика

публікацій, інтересів користувача, предметів та інше, що зробить рекомендації більш специфічними і точними;

- впровадження методів машинного навчання для прогнозування й аналізу очікуваної продуктивності на основі поточного прогресу та історичних даних, що допоможе передбачити труднощі та стратегії їх усунення;
- використання засобів машинного навчання для аналізу складних закономірностей, контексту та покращення точності рекомендацій;
- розробка інтерфейсу в інтелектуалізованій системі аналізу діяльності та рейтингування викладачів для взаємодії користувачів, отримання рекомендацій, можливість їх оцінювання користувачами.

СПИСОК ЛІТЕРАТУРИ

1. Рейтингова система Західноукраїнського національного університету. URL: <https://rating.wunu.edu.ua/> (дата звернення: 24.03.2025).
2. Intelligent Analyzing Module in the Academic Staff Performance Appraisal System / A. Pukas et al. *2023 12th International Conference on Advanced Computer Information Technologies (ACIT)*. 2023. P. 423–429.
3. Automated Reporting Module in the Academic Staff Performance Appraisal System / A. Pukas et al. *2024 14th International Conference on Advanced Computer Information Technologies (ACIT)*. 2024. P. 423–429.
4. Модуль збору даних для автоматичного заповнення звіту в інтелектуалізованій системі аналізу діяльності та рейтингування викладачів / А. В. Пукас та ін. *Матеріали зимової школи-семінару молодих вчених і студентів «Комп'ютерні інформаційні технології» CIT'2024*. Тернопіль: ЗУНУ, 2024. С. 69–71.
5. A user-knowledge vector space reconstruction model for the expert knowledge recommendation system / L. Gao et al. *Information Sciences*. 2023. Vol. 632. P. 358–377. URL: <https://doi.org/10.1016/j.ins.2023.03.025>.
6. Nuar A. N. A., Sen S. C. Examining the Trend of Research on Big Data Architecture: Bibliometric Analysis using Scopus Database. *Procedia Computer Science*. 2024. Vol. 234. P. 172–179. URL: <https://doi.org/10.1016/j.procs.2024.04.010>.
7. Arpita C., Monalisa S., Debasis S. SHARE: Designing multiple criteria-based personalized research paper recommendation system. *Information Sciences*. 2022. Vol. 617. P. 41–64. URL: <https://doi.org/10.1016/j.ins.2022.09.064>.
8. Recommender System in Academic Choices of Higher Education: A Systematic Review / N. Kamal et al. *IEEE Access*. 2024. Vol. 12. P. 35475–35501. URL: <https://doi.org/10.1109/ACCESS.2024.3368058>.
9. Hodovychenko M., Gorbatenko A. Recommender systems: models, challenges and opportunities. *Herald of Advanced Information Technology*. 2023. Vol. 6. P. 308–319. URL: <https://doi.org/10.15276/hait.06.2023.20>.
10. Cypher Basics I. Neo4j Developer. URL: <https://neo4j.com/developer/cypher-basics-i/> (дата звернення: 12.03.2025).
11. Aldwairi M., Jarrah M., Mahasneh N., Al-khateeb B. Graph-based data management system for efficient information storage, retrieval and processing. *Information Processing & Management*. 2023. Vol. 60, Issue 2. URL: <https://doi.org/10.1016/j.ipm.2022.103165>.
12. Петрина В. В., Дорошенко А. В. Ефективність застосування методів класифікації для задач інтелектуального аналізу великих даних. *Scientific Bulletin of UNFU*. 2024. Т. 34, №5. С. 119–128. URL: <https://doi.org/10.36930/40340516>.
13. E-Learning Course Recommender System Using Collaborative Filtering Models / K. K. Jena et al. *Electronics*. 2022. Vol. 12 (1). Article ID: 157. URL: <https://doi.org/10.3390/electronics12010157>.
14. Ondiek B., Mburu L., Njenga S. Machine Learning Applications for Recommender Systems in Higher Education: A Systematic Review. *International Journal Of Applied Research And Technology*. 2023. Vol. 8. URL: https://www.researchgate.net/publication/378970216_Machine_Learning_Applications_for_Recommender_Systems_in_Higher_Education_A_Systematic_Review.
15. Jain G., Mahara T., Sharma S. C. Performance Evaluation of Time-based Recommendation System in Collaborative Filtering Technique. *Procedia Computer Science*. 2023. Vol. 218. P. 1834–1844. URL: <https://doi.org/10.1016/j.procs.2023.01.161>.
16. Falah Z., Suryawan F. Recommendation System to Propose Final Project Supervisors using Cosine Similarity Matrix. *Khazanah Informatika: Jurnal Ilmu Komputer dan Informatika*. 2022. Vol. 8. URL: <https://doi.org/10.23917/khif.v8i2.16235>.
17. A Review On Collaborative Filtering Using KNN Algorithm / A. Sharma et al. *OPJU International Technology Conference on Emerging Technologies for Sustainable Development (OTCON)*. 2023. P. 1–6. URL: <https://doi.org/10.1109/OTCON56053.2023.10113985>.

Стаття надійшла до редакції 21.05.2025.

Стаття пройшла рецензування 27.05.2025.

Пукас Андрій Васильович – д-р техн. наук, професор, завідувач кафедри комп'ютерних наук,
e-mail: apu@wunu.edu.ua.

Сімак Андрій Юрійович – аспірант, викладач кафедри комп'ютерних наук,
e-mail: a.simak@wunu.edu.ua.

Західноукраїнський національний університет.