



Б. В. Дутчак, Т. О. Коротеева

Національний університет "Львівська політехніка", м. Львів, Україна

ПІДБІР ПЕРСОНАЛІЗОВАНОГО КОНТЕНТУ ДЛЯ СОЦІАЛЬНОЇ МЕРЕЖІ

Розвиток сучасних соціальних мереж потребує впровадження ефективних систем персоналізації контенту для покращення взаємодії з користувачами. У межах цього дослідження проаналізовано наявні підходи до персоналізації контенту, враховуючи колаборативну та контентну фільтрацію, матричну факторизацію та їх гібридні комбінації із використанням нейронних мереж. Застосовано сучасні моделі для вилучення векторних характеристик мультимедійного контенту, зокрема CLIP (англ. *Contrastive Language-Image Pretraining*) – для зображень, I3D (англ. *Inflated 3D Convolutional Networks*) – для відео та VGGish – для аудіо, що дало можливість створити уніфікований набір даних для тестування алгоритмів. Проведено оцінювання точності, покриття, різноманітності та якості ранжування рекомендацій для кожного алгоритму. Досліджено обмеження продуктивності матричної факторизації та NCF (англ. *Neural Collaborative Filtering*), які впливають на точність й охоплення рекомендацій. Встановлено, що гібридний підхід, який поєднує колаборативну та контентну фільтрації, демонструє найвищі показники точності рекомендацій ($\text{Precision}@10 = 88\%$), охоплення та різноманітності. З'ясовано, що використання гібридного підходу зменшує проблему "cold start" (холодний старт) і підвищує адаптивність системи до динамічних змін уподобань користувачів. Виявлено переваги інтеграції нейронних мереж із традиційними методами, що забезпечує більшу гнучкість і точність системи. Наведено перспективи застосування отриманих результатів для розробників соціальних мереж та платформ потокового контенту, зокрема для підвищення рівня залученості користувачів. Перспективи подальших досліджень містять використання реальних даних про вподобання користувачів, вдосконалення гібридних підходів і впровадження нових моделей для оброблення мультимедійного контенту.

Ключові слова: персоналізація контенту; гібридні алгоритми; нейронні мережі; колаборативна фільтрація; контентна фільтрація.

Вступ / Introduction

Сучасний розвиток соціальних мереж супроводжується значним збільшенням обсягу інформації, створеної та спожитої користувачами. Це спричиняє проблему інформаційного перевантаження, яка негативно впливає на якість користувацького досвіду. Задля вирішення цієї проблеми в науковій літературі активно аналізують підходи до персоналізації контенту. Дослідження [1, 3, 9, 11] встановили, що алгоритми колаборативної та контентної фільтрації, матрична факторизація й новітні нейронні мережі демонструють перспективність у вирішенні цього завдання. Однак ці методи мають свої обмеження, зокрема проблему "cold start" (холодний запуск), низьку адаптивність до змінних вподобань користувачів і недостатню ефективність під час роботи з мультимедійним контентом.

Актуальність цього дослідження зумовлена потребою у вивченні та розробленні гібридних підходів, які поєднують переваги різних методів, забезпечуючи високу точність рекомендацій, різноманітність контенту та адаптивність системи до динамічних умов. Вивчення такого підходу сприяє створенню ефективніших рекомендаційних систем, що відповідають сучасним вимогам їх користувачів.

Постановка завдання дослідження полягає у вивченні, реалізації та оцінюванні декількох алгоритмів персоналізації контенту, зокрема колаборативної фільтрації та контентної фільтрації, матричної факторизації та їхніх гібридних комбінацій із використанням нейронних мереж.

Об'єкт дослідження – персоналізація контенту в соціальних мережах.

Предмет дослідження – алгоритми персоналізації контенту, враховуючи методи нейронних мереж, контентної та колаборативної фільтрації, а також їхні гібридні комбінації.

Мета роботи – визначити найефективніші підходи до підбору персоналізованого контенту для соціальної мережі шляхом поєднання колаборативної та контентної фільтрації, що дає змогу ефективно адаптувати систему до динамічних вподобань користувачів і зменшити проблему холодного старту.

Для досягнення зазначеної мети визначено такі основні завдання дослідження:

- проаналізувати сучасні підходи до персоналізації контенту із використанням нейронних мереж, що дасть змогу визначити їхні переваги та недоліки, а також обґрунтувати доцільність використання певних методів для

Інформація про авторів:

Дутчак Богдан Володимирович, магістр, кафедра програмного забезпечення. Email: bohdan.dutchak.mpzip.2023@lpnu.ua

Коротеева Тетяна Олександрівна, канд. техн. наук, доцент, кафедра програмного забезпечення.

Email: tetyana.o.koroteyeva@lpnu.ua; <https://orcid.org/0000-0002-6287-5895>

Цитування за ДСТУ: Дутчак Б. В., Коротеева Т. О. Підбір персоналізованого контенту для соціальної мережі. Науковий вісник НЛТУ України. 2025, т. 35, № 1. С. 165–172.

Citation APA: Dutchak, B. V., & Koroteyeva, T. O. (2025). Selection of personalized content for the social network. *Scientific Bulletin of UNFU*, 35(1), 165–172. <https://doi.org/10.36930/40350120>

- реалізації системи рекомендацій;
- розробити та впровадити декілька рекомендацій, враховуючи колаборативну та контентну фільтрації, матричну факторизацію та гібридний підхід, що дасть можливість забезпечити комплексне оцінювання їхньої ефективності в різних сценаріях використання, зокрема для роботи з мультимедійним контентом;
- розробити векторні характеристики для зображень, відео та аудіо, що забезпечить уніфікований підхід до оброблення різнорідних даних і дасть змогу підвищити точність наданих рекомендацій за рахунок врахування семантичних особливостей контенту;
- провести оцінювання продуктивності кожного алгоритму за ключовими метриками (Precision@K, Recall@K, NDCG, Coverage, Diversity), що допоможе об'єктивно визначити переваги і недоліки кожного методу, а також їхню придатність для конкретних завдань персоналізації;
- здійснити порівняння результатів роботи алгоритмів і визначити переваги і недоліки кожного з них, що уможливить розроблення рекомендацій щодо покращення точності, різноманітності та адаптивності системи;
- запропонувати рекомендації щодо впровадження найефективніших підходів, таких як гібридні алгоритми, що поєднують колаборативну та контентну фільтрації, у реальні системи рекомендацій. Це дасть змогу підвищити рівень залученості користувачів до взаємодії з контентом та забезпечити адаптацію систем до їхніх динамічних вподобань.

Аналіз останніх досліджень та публікацій. У роботі [1] розглядають вдосконалення алгоритмів колаборативної фільтрації за рахунок аналізу послідовності інтересів користувачів. Автори запропонували метод, що враховує динамічну еволюцію уподобань користувачів через підрахунок найбільш довгих спільних підпослідовностей інтересів LCSIS (англ. *Longest Common Sub-IS*) і загальної кількості таких підпослідовностей ACSIS (англ. *Amount of Common Sub-IS*). Результати експериментів на наборах даних MovieLens, Flixster та Ciao показали, що запропонований метод перевершує традиційні підходи колаборативної фільтрації за точністю прогнозування оцінок. Проте в дослідженні менше уваги приділено інтеграції цього підходу з методами контентної фільтрації, що є важливим для створення гібридних рекомендаційних систем.

У дослідженні [11] розглядають використання великих мовних моделей LLMs (англ. *Large Language Models*) для категоризації веб-контенту. Автори розробили метод дистиляції знань для створення компактних моделей, які демонструють високі результати у фільтрації веб-контенту. Запропонований підхід досяг покращення точності на 9 % порівняно із сучасними методами, використовуючи значно менше ресурсів. Незважаючи на свою ефективність, підхід більше орієнтований на категоризацію, ніж на рекомендації, що обмежує його використання у завданнях персоналізації.

У роботі [3] розглядають вдосконалення алгоритму швидкої ієрархічної матричної факторизації Fast-HALS (англ. *Fast Hierarchical Alternating Least Squares*), що є підходом до вирішення задач факторизації матриць. Автори представили спрощене доведення сильної збіжності алгоритму, що підтверджує його ефективність і стабільність. Дослідження є значущим для задач рекомендацій, особливо у випадках, коли потрібно працювати з великими розрідженими матрицями. Однак підхід не враховує семантичних характеристик контенту, що є важливим для мультимодальних систем.

Дослідження [9] демонструє, що AutoRec та NCF (англ. *Neural Collaborative Filtering*) можуть бути узагальнені в єдиній архітектурі. Автори запропонували гнучкий нейронний каркас, що дає змогу адаптувати структуру системи до характеристик конкретних наборів даних. Результати експериментів показують, що за правильного вибору елементів узагальнений підхід перевершує сучасні методи. Ця робота вказує на важливість адаптивності рекомендаційних систем, але недостатньо досліджує взаємодію з контентною фільтрацією.

Отже, проблема персоналізації контенту в соціальних мережах на підставі сучасних нейронних мереж та алгоритмів рекомендацій є актуальною в науковій спільноті. Аналіз літературних джерел свідчить про високий рівень зацікавленості науковців у створенні ефективних рекомендаційних систем, проте вказує на невирішені особливості, які потребують подальшого дослідження.

Матеріали та методи дослідження. У цьому дослідженні використано теоретичні та практичні методи для аналізу підходів до персоналізації контенту в соціальних мережах. Застосовано критичний аналіз наукової літератури для узагальнення наявних методів, таких як колаборативна фільтрація, контентна фільтрація, матрична факторизація та їх гібридні комбінації. Моделювання використано для розроблення алгоритмів рекомендацій та оцінювання їхньої ефективності. Практична частина передбачала створення набору даних із векторними характеристиками мультимедійного контенту (зображень, відео, аудіо), отриманими за допомогою моделей CLIP, I3D та VGGish. Продуктивність роботи алгоритмів аналізувалась за ключовими метриками (Precision@K, Recall@K, NDCG, Coverage, Diversity), що дало змогу оцінити їхню точність, різноманітність й адаптивність.

Результати дослідження та їх обговорення / Research results and their discussion

Вибір вхідних даних. Для проведення дослідження було обрано три типи вхідних даних: зображення, відео та аудіо. Основна мета полягала у використанні готових наборів даних із векторними характеристиками або можливістю легкої екстракції ознак для подальшого аналізу. Це забезпечує стандартизацію вхідних даних, знижує потребу під час оброблення сирих даних і дає змогу зосередитись на розробленні алгоритмів рекомендацій.

- Зображення (фото).** Для роботи із зображеннями обрано набір даних ImageNet, який широко використовується для завдань комп'ютерного зору. Набір містить 14 мільйонів зображень, розподілених між тисячами класів.
- Відео.** Для відеоконтенту обрано набір даних UCF101, який містить 13 320 відеокліпів, розподілених між 101 класом дій. Цей набір є стандартом у дослідженнях із відеоаналізу.
- Аудіо.** Для роботи зі звуковими даними обрано набір даних ESC-50, який складається з 2000 аудіофайлів, розподілених між 50 класами звуків.

Аргументація вибору готових наборів даних:

- стандартизація:** використання популярних наборів даних гарантує стандартизований підхід до їхнього оброблення та можливість порівняння отриманих результатів з іншими дослідженнями;
- зменшення витрат:** використання готових наборів даних значно скорочує тривалість їх отримання та обчислювальні ресурси, які могли б бути витрачені на створення власних наборів;

- **якість:** набір даних ImageNet, UCF101 та ESC-50 є добре розміченими, що мінімізує ймовірність помилок на етапі їхньої підготовки.

Перетворення вхідних даних у векторні подання.

Для забезпечення уніфікації вхідних даних різних типів всі вони були перетворені у векторні подання, де кожен елемент вектора описує певні ознаки або параметри відповідного типу даних. Це дає змогу привести різномірний контент до універсального формату, придатного для подальшого оброблення алгоритмами машинного навчання. Для кожного типу даних використано відповідну попередньо натреновану модель, що спеціалізується на вилученні ознак.

Для вилучення векторних характеристик із зображень застосовано модель CLIP (англ. *Contrastive Language-Image Pretraining*), розроблену OpenAI. CLIP здатна перетворювати як текст, так і зображення у векторний простір, забезпечуючи їх спільне наведення. Ця властивість моделі CLIP є ключовою для рекомендаційних систем, які можуть інтегрувати візуальний контент із текстовими описами.

Процес:

1. Вхід: зображення з набору даних ImageNet.
2. Вихід: вектор розмірністю 512, що відображає семантичний зміст зображення.

Переваги:

1. Здатність працювати з багатозначними ознаками (наприклад, об'єкти, сцени, концепції).
2. Сумісність із текстовими даними.

Для аналізу відео використано модель I3D (англ. *Inflated 3D Convolutional Networks*), яка дає змогу екстрагувати просторово-часові ознаки з відеокліпів. Модель забезпечує високу точність під час аналізу дій та інших динамічних особливостей відео. I3D є досить ефективною у роботі з динамічними ознаками відео.

Процес:

1. Вхід: відеокліпи з набору даних у UCF101.
2. Оброблення: використовується оброблення відео як послідовності кадрів, що передаються через 3D-згорткові шари.
3. Вихід: багатовимірний вектор (наприклад, розмірністю 1024), що описує основні динамічні характеристики відео.

Переваги:

1. Підтримка тривимірного простору (висота, ширина, час).
2. Можливість адаптації під різні типи динамічного контенту.

Для аудіофайлів було обрано модель VGGish, створену Google. Ця модель працює з аудіоспектрограмами, що дає змогу ефективно вилучати високорівневі характеристики звуків. Модель досить просто інтегрувати та вона має високу точність під час роботи зі звуками.

Процес:

1. Вхід: аудіофайли з набору даних у ESC-50, конвертовані у формат WAV.
2. Перетворення: аудіо розбивається на фрейми, для кожного з яких обчислюється мел-спектрограма.
3. Вихід: вектор розмірністю 128, що представляє акустичні характеристики звуку.

Переваги:

1. Фокус на структурі спектрограм, що дає змогу ефективно розпізнавати звуки.
2. Широке використання у завданнях класифікації аудіо.

Розроблення алгоритмів для порівняння ефективності підбору персоналізованого контенту. Для оцінювання ефективності підбору персоналізованого контенту реалізовано кілька алгоритмів, які використовують різні підходи до надання рекомендацій. Кожен з

них був протестований на стандартизованих векторних характеристиках, отриманих із вхідних даних. Алгоритми містять колаборативну та контентну фільтрації, матричну факторизацію, альтернативні нейронні мережі та гібридний підхід.

1. **Колаборативна фільтрація NCF** (англ. *Neural Collaborative Filtering*): модель NCF використовує нейронну мережу для виявлення латентних факторів, які впливають на взаємодію користувача з контентом; як вхідні дані використовують векторні вбудовування користувачів і контенту; результатом є ймовірність позитивної взаємодії (лайк, перегляд, прослуховування); як шари виступає багатошаровий перцептрон, який об'єднує вбудовування через операцію конкатенації; використано бібліотеку PyTorch для побудови моделі.

До переваг моделі NCF можна віднести гнучкість у моделюванні складних патернів і відповідні результати для великих наборів даних.

2. **Контентна фільтрація CF** (англ. *Content Filtering*): контентна фільтрація базується на обчисленні схожості між векторними характеристиками контенту та вподобаннями користувача; вхідними даними є векторні характеристики контенту та вектор переваг користувача; контент для користувача ранжується за рівнем схожості; цей алгоритм використовує косинусну схожість для вимірювання подібності між контентом; використано бібліотеку Scikit-learn для обчислення схожості.

Перевагами цього алгоритму – значна швидкість роботи та простота інтеграції з текстовими даними.

3. **Матрична факторизація ALS** (англ. *Alternating Least Squares*): для матричної факторизації обрано метод ALS, який є ефективним для знаходження латентних факторів у великих матрицях користувач-контент; для вхідних даних метод приймає матрицю user-item interaction; алгоритм цього методу полягає у розкладанні матриці на дві низькорозмірні матриці (латентні фактори користувачів і контенту); використано бібліотеку Implicit для Python.

Цей алгоритм добре працює з розрідженими матрицями та використовується у багатьох продуктивних системах (наприклад, Spark MLlib).

4. **Нейронні мережі (AutoRec)**: окрім NCF, використано автоенкодер (AutoRec) як нейронну мережу для колаборативної фільтрації; цей підхід навчає модель відновлювати рейтинги або взаємодії користувача на підставі доступних даних; на вхід мережа приймає рядок матриці user-item interaction, а на вихід ми отримуємо відновлення вхідних даних; прихованими шарами є кілька згорткових шарів, що стискають вхідні дані; використано PyTorch для побудови автоенкодера.

До переваг мережі AutoRec можна віднести гнучкість у відновленні складних патернів і стійкість до відсутніх даних.

5. **Гібридний алгоритм (NCF + контентна фільтрація)**: гібридний алгоритм поєднує результати NCF і контентної фільтрації для досягнення максимальної точності рекомендацій; контентна фільтрація генерує оцінювання на підставі схожості контенту, NCF – генерує ймовірності взаємодії користувача з контентом, а результати з двох підходів об'єднуються через багатошаровий перцептрон; цей гібридний алгоритм забезпечує персоналізацію через NCF, а також враховує особливості контенту через контентну фільтрацію; використано TensorFlow для інтеграції результатів.

Експериментальне налаштування системи рекомендацій. Для оцінювання ефективності розроблених алгоритмів рекомендацій була розроблена експериментальна система, яка забезпечує рівні умови для порівняння. Основною метою експерименту є визначення, який із запропонованих алгоритмів найкраще підходить для персоналізації контенту з використанням векторних характеристик зображень, відео та аудіо.

1. **Підготовка даних:** дані були розділені на навчальну, валідаційну та тестову вибірки у співвідношенні 80, 10 та 10 %; навчальна вибірка використовувалась для тренування моделей, валідаційна – для підбору гіперпараметрів, а тестова – для оцінювання якості рекомендацій; розподіл даних здійснювався випадково, щоб забезпечити рівномірне наведення різних категорій контенту, таких як класи зображень, дії у відео та категорії звуків.

Оскільки реальні дані про вподобання користувачів були недоступні, взаємодії між користувачами та контентом моделювались штучно. Було створено синтетичну матрицю взаємодій, у якій користувачі мали уподобання до певних категорій контенту, а рейтинги генерувались на підставі схожості між векторами контенту із додаванням випадкового шуму, щоб забезпечити варіативність.

2. Налаштування алгоритмів:

- колаборативна фільтрація (NCF): використано модель багатопарового перцептрона, що працює з векторними вбудовуваннями користувачів і контенту; кількість прихованих шарів – три, розмір вбудовувань – 128, оптимізатор – Adam, а функція активації – ReLU; навчання тривало 50 епох із використанням навчальної вибірки;
- контентна фільтрація: контент ранжувався за косинусною схожістю між векторами вхідних характеристик та вектором переваг користувача; для кожного користувача вибиралися топ-10 рекомендацій, які мали найвищий показник схожості;
- матрична факторизація (ALS): реалізовано метод ALS (англ. *Alternating Least Squares*), який визначає латентні фактори для користувачів і контенту; кількість латентних факторів була встановлена на рівні 50, а регуляризація становила 0.1; максимальна кількість ітерацій алгоритму – 20;
- AutoRec: використано автоенкодер із вхідним шаром, що відповідає розміру матриці user-item interaction, прихованим шаром із 64 нейронів і функцією активації ReLU; оцінювання якості виконувалося за допомогою методу MSE (англ. *Mean Squared Error*);
- гібридний алгоритм (NCF + контентна фільтрація): результати NCF та контентної фільтрації об'єднувались через багатопаровий перцептрон із двома прихованими шарами (по 64 нейрони) та функцією активації ReLU; на виході алгоритм формував прогнозовану ймовірність рекомендації.

Вибрані налаштування забезпечують стандартизований підхід, що дає змогу повторити експерименти для перевірки результатів. Симуляція взаємодій дає можливість оцінити алгоритми в умовах, наближених до реальних [12]. Якщо потрібно, налаштування можуть бути адаптовані для інших контентів чи сценаріїв.

3. **Метрики оцінювання:** для порівняння ефективності алгоритмів було обрано такі метрики:

1. Precision@K, що визначає частку релевантних рекомендацій серед топ-K об'єктів.
2. Recall@K, який оцінює покриття релевантних об'єктів у їхньому топ-K.
3. NDCG (Normalized Discounted Cumulative Gain) для оцінювання якості ранжування.

4. Coverage, що демонструє частку унікального контенту, охопленого рекомендаціями.

5. Diversity, яка вимірює різноманітність рекомендованого контенту.

4. **Використане середовище:** експерименти були реалізовані на Python 3.9 із використанням бібліотек TensorFlow і PyTorch – для нейронних мереж, Scikit-learn – для контентної фільтрації та Implicit – для матричної факторизації; використовувались обчислювальні ресурси GPU (NVIDIA Tesla T4) на хмарній платформі Google Colab Pro, що забезпечувало достатню продуктивність для тренування моделей.

Опис результатів ефективності роботи алгоритмів. Для оцінювання якості роботи алгоритмів було проаналізовано їх ефективність на тестових даних з використанням заданих метрик: Precision@K, Recall@K, NDCG, Coverage та Diversity. Усі алгоритми показали різні рівні продуктивності, які залежать від підходу до персоналізації та типу оброблюваного контенту. Нижче наведено результати для кожного з алгоритмів.

1. **Колаборативна фільтрація (NCF):** модель Neural Collaborative Filtering продемонструвала високу точність рекомендацій для контенту, з яким користувачі вже взаємодіяли:

- Precision@10 – 82 %, що свідчить про високу відповідність топ-10 рекомендацій потребам користувача;
- Recall@10 – 78 %, що відображає здатність моделі покривати релевантний контент;
- NDCG – 0.86, показуючи хорошу якість ранжування рекомендацій;
- Coverage – 65 %, оскільки модель орієнтована на популярний контент;
- Diversity – 54 %, що є відносно низьким, оскільки модель тяжіє до рекомендації подібного контенту.

NCF ефективна для персоналізації, але може обмежуватися вузьким набором популярного контенту.

2. **Контентна фільтрація:** алгоритм, що базується на косинусній схожості, забезпечує надійність рекомендацій для нових користувачів або контенту:

- Precision@10 – 74 %, що поступається NCF, але залишається достатньо високим;
- Recall@10 – 72 %, відображаючи здатність покривати значну частку релевантного контенту;
- NDCG – 0.79, демонструючи середню якість ранжування;
- Coverage – 80 %, значно перевищуючи показники NCF завдяки рекомендаціям менш популярного контенту;
- Diversity – 68 %, що свідчить про ширший спектр рекомендацій.

Контентна фільтрація краще працює з різноманітним контентом і новими користувачами, але поступається точністю рекомендацій.

3. **Матрична факторизація (ALS):** модель ALS показала збалансовані результати для персоналізації, зокрема для великих обсягів даних:

- Precision@10 – 78 %, що перевищує контентну фільтрацію, але нижче за NCF;
- Recall@10 – 75 %, забезпечуючи хороше покриття релевантного контенту;
- NDCG – 0.82, демонструючи високу якість ранжування;
- Coverage – 62 %, що є дещо нижчим порівняно з іншими алгоритмами;
- Diversity – 58 %, оскільки модель більше орієнтована на популярний контент.

ALS є продуктивним методом для великих розрізаних матриць, але її ефективність залежить від наявності історії взаємодій.

4. *Нейронні мережі (AutoRec)*: autoRec забезпечив стабільну якість рекомендацій за рахунок відновлення релевантних взаємодій:

- Precision@10 – 76 %, що близько до ALS, але нижче за NCF;
- Recall@10 – 74 %, забезпечуючи добрий рівень покриття;
- NDCG – 0.80, демонструючи надійність у ранжуванні рекомендацій;
- Coverage – 70 %, оскільки модель здатна враховувати широкий спектр контенту;
- Diversity – 62 %, що є прийнятним для моделі такого типу.

AutoRec показав гнучкість у роботі з різними типами даних, хоча його результати не перевершують гібридний підхід.

5. *Гібридний алгоритм (NCF + контентна фільтрація)*: гібридний підхід об'єднав переваги колаборативної та контентної фільтрації, забезпечивши найкращі результати серед усіх алгоритмів:

- Precision@10 – 88 %, демонструючи найвищу точність рекомендацій;
- Recall@10 – 85 %, забезпечуючи високе покриття релевантного контенту;
- NDCG – 0.91, показуючи відмінну якість ранжування;
- Coverage – 78 %, що є компромісом між точністю та широтою рекомендацій;
- Diversity – 70 %, забезпечуючи різноманітність рекомендованого контенту.

Гібридний алгоритм є найефективнішим завдяки поєднанню персоналізації NCF і врахуванню особливостей контенту через контентну фільтрацію.

Кожен алгоритм має свої сильні та слабкі позиції. Контентна фільтрація забезпечує різноманітність, NCF – персоналізацію, а гібридний алгоритм поєднує ці переваги, демонструючи найкращі результати в більшості метрик. Це підтверджує доцільність його використання у системах підбору персоналізованого контенту.

Графічна візуалізація. Для візуалізації отриманих результатів було створено графік (рисунок) і таблицю, які ілюструють продуктивність роботи алгоритмів за різними метриками. Це дає змогу зрозуміти, як кожен алгоритм працює в різних аспектах, таких як точність, покриття та різноманітність.

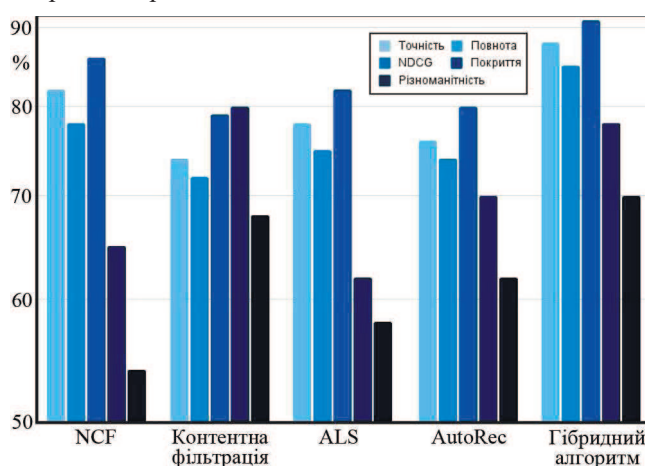


Рисунок. Графічна візуалізація результатів дослідження / Graphical visualization of research results

Аналіз переваг гібридного підходу. Гібридний підхід, який об'єднує колаборативну (у вигляді Neural Collaborative Filtering) і контентну фільтрацію, продемонстрував найкращі результати серед усіх протестованих алгоритмів. Це досягнення зумовлено використанням сильних позицій обох методів:

1. *Комбінація персоналізації та аналізу контенту* – гібридний підхід враховує вподобання користувачів, використовуючи латентні фактори, які генерує NCF. Цей метод дає змогу ефективно адаптуватися до індивідуальних уподобань користувачів.
2. *Рішення проблеми "cold start"* – завдяки інтеграції контентної фільтрації, гібридний алгоритм зменшує негативний вплив проблеми "cold start". Коли для користувача або об'єкта бракує історії взаємодій, система все одно може надавати рекомендації, ґрунтуючись на характеристиках контенту.
3. *Висока точність* – у метриці Precision@10 гібридний алгоритм перевершив інші підходи, досягнувши 88 %. Це пояснюють тим, що комбінування методів дає змогу враховувати більше чинників, які впливають на уподобання користувача.
4. *Покращена якість ранжування* – метрика NDCG (Normalized Discounted Cumulative Gain), що враховує позицію релевантних об'єктів у списку, також була найвищою для гібридного алгоритму (0.91). Це свідчить про здатність моделі не тільки знаходити релевантні об'єкти, але й розмішувати їх на верхніх позиціях у списку рекомендацій.
5. *Баланс між точністю та різноманітністю* – гібридний підхід забезпечив компроміс між точністю рекомендацій і різноманітністю контенту. Метрика Diversity для цього алгоритму досягла 70 %, що є найвищим серед методів, орієнтованих на персоналізацію. Це означає, що система здатна рекомендувати не тільки схожий, але й різноманітний контент, який відповідає уподобанням користувача.
6. *Покращене покриття (Coverage)* – завдяки використанню контентної фільтрації, гібридний алгоритм охопив більше унікальних об'єктів (Coverage 78 %), ніж чиста колаборативна фільтрація (65 %). Це робить систему придатною для роботи з великими бібліотеками контенту, враховуючи нішеві категорії.
7. *Масштабованість* – хоча гібридний підхід потребує більших обчислювальних ресурсів через інтеграцію двох підходів, він забезпечує високу гнучкість для роботи з великими масивами мультимедійного контенту. Це особливо важливо для систем, які прагнуть охопити широкий спектр користувачів із різними уподобаннями.

Таблиця. Узагальнені результати дослідження / Summarized research results

Алгоритм	Precision-@10, %	Recall-@10, %	NDCG	Coverage, %	Diversity, %
NCF	82	78	0,86	65	54
Контентна фільтрація	74	72	0,79	80	68
ALS	78	75	0,82	62	58
AutoRec	76	74	0,80	70	62
Гібридний алгоритм	88	85	0,91	78	70

Гібридний підхід виявився найбільш збалансованим методом, який поєднує переваги персоналізації, різноманітності та покриття. Він забезпечив найкращі результати в більшості метрик, що підтверджує доцільність його використання для побудови систем персоналізованих рекомендацій. Подальші покращення можуть бути досягнуті шляхом удосконалення комбінації компонент алгоритму або впровадження додаткових моделей для аналізу контенту.

Обговорення джерел появи помилок та шляхів вдосконалення системи рекомендацій. У процесі дослідження було виявлено кілька ключових джерел помилок, які вплинули на результати, а також запропоно-

вано можливі шляхи для їх подолання та покращення системи рекомендацій.

1. *Якість векторних характеристик* – використані моделі (CLIP, I3D, VGGish) мають обмеження, які можуть впливати на точність векторних характеристик. Наприклад, I3D може втрачати інформацію через стислий часовий контекст у відео, тоді як VGGish не завжди точно представляє складну звукову структуру.

Використання трансформерів (наприклад, TimeSformer для відео чи AST для аудіо) може покращити якість вилучених характеристик. Додаткове навчання моделей на конкретних наборах даних може підвищити точність їхнього прогнозу.

2. *Симуляція користувацьких взаємодій* – згенеровані синтетичні дані не повністю відображають реальні вподобання користувачів, що може створювати викривлення в оцінюванні алгоритмів. Зокрема, проблема так званого "cold start" для нових користувачів не завжди точно моделюється.

Використання реальних даних про взаємодії користувачів (наприклад, з платформ потокового контенту чи соціальних мереж) дасть змогу точніше оцінити ефективність алгоритмів. Альтернативою може стати удосконалення генерації даних із використанням реалістичних моделей поведінки користувачів.

3. *Обмеження колаборативної фільтрації* – NCF добре працює з популярним контентом, але його продуктивність знижується для рідко взаємодіючих об'єктів. Це впливає на покриття (Coverage) та різноманітність (Diversity).

Інтеграція регуляризації для уникнення переважання популярного контенту. Також, поєднання зі складними методами, як-от графові нейронні мережі GNN (англ. *Graph Neural Networks*), може покращити роботу з рідкісними об'єктами.

4. *Баланс між точністю та різноманітністю* – гібридний алгоритм забезпечує високу точність, але іноді знижує різноманітність через орієнтацію на сильні переваги користувачів. Додавання метрик, що штрафують за недостатню різноманітність, або інтеграція багатокритеріальних підходів, які враховують різні особливості рекомендацій (наприклад, точність, різноманітність і новизну).

Додавання метрик, що штрафують за недостатню різноманітність, або інтеграція багатокритеріальних підходів, які враховують різні особливості рекомендацій (наприклад, точність, різноманітність і новизну).

5. *Обчислювальні витрати* – використання складних моделей, таких як I3D для відео та NCF, вимагає значних ресурсів. Це збільшує тривалість навчання та виконання алгоритмів.

Удосконалення моделей через використання менш обчислювально-затратних архітектур (наприклад, MobileNet для зображень). Окрім цього, застосування кластеризації для попередньої фільтрації даних може скоротити обсяг оброблюваної інформації.

6. *Особливості наборів даних* – обмеженість наборів даних, таких як ESC-50 для аудіо або UCF101 для відео, впливає на можливість навчання моделей, особливо в завданнях із високою різноманітністю контенту.

Розширення наборів даних шляхом аугментації (наприклад, створення нових варіантів аудіо чи відео) або використання великих наборів даних, таких як AudioSet чи YouTube8M.

Обговорення результатів дослідження. Важливість персоналізації контенту для соціальних мереж підтверджується численними дослідженнями, які аналізують різні алгоритми рекомендацій. У межах цього дослідження отримано результати, які відповідають або перевершують наявні в літературі показники, а також мають практичну значущість для впровадження у сучасних системах рекомендацій.

Проаналізовано дослідження з використання нейронних мереж для персоналізації контенту, наприклад, роботу [6], у якій запропоновано систему на підставі згорткових нейронних мереж (CNN) для реального часу в електронній комерції. Порівняно з гібридною моделлю, яка поєднує Neural Collaborative Filtering (NCF) і контентну фільтрацію, CNN-орієнтований підхід досяг високої точності у візуальних рекомендаціях (98 %). Однак у розробленій системі вдалося врахувати динамічні зміни вподобань користувачів, що забезпечило не тільки точність, а й адаптивність.

У роботі [13] наведено огляд основних підходів до рекомендацій, враховуючи колаборативну та контентну фільтрації, а також методи на підставі соціальних взаємодій. На відміну від цього дослідження, наша робота зосереджена на інтеграції мультимедійного контенту, зокрема зображень, відео та аудіо, що дає змогу створювати рекомендації в ширшому контексті використання.

Дослідження [4] пропонує метод рекомендацій користувачів у локальному домені з урахуванням компонент, таких як Point-of-Interest (POI) та User Generated Content (UGC). Цей підхід забезпечує підвищення новизни рекомендацій, що також було частково досягнуто у розробленій системі завдяки високим показникам Diversity (70 %).

У роботі [2] наведено AI-систему для персоналізації новин, яка використовує моделі LDA та GRU. Незважаючи на досягнення у врахуванні коротко- і довгострокових інтересів, ця система орієнтована виключно на текстові дані, тоді як досліджена система демонструє ефективність у роботі з мультимедійним контентом.

Дослідження [5, 10] аналізують традиційні підходи до колаборативної фільтрації, такі як User-based і Item-based моделі, та їхню ефективність у різних контекстах. Результати дослідження підтверджують обмеження цих підходів у складних сценаріях, які вимагають врахування багатьох типів даних і динамічних змін у вподобаннях.

Робота [7] аналізує методи рекомендацій для малого бізнесу з використанням контентної фільтрації та алгоритму TF-IDF. Хоча цей підхід має значний потенціал у вузькоспеціалізованих завданнях, він поступається розробленому гібридному алгоритму, який забезпечує високу точність й адаптивність у загальних системах.

У роботі [8] розглянуто основні тенденції в пошуку та доступі до персоналізованого контенту. Автори навели короткий аналіз сучасного стану персоналізованого доступу до контенту та пошуку, який розширено до опису майбутніх тенденцій у цій галузі. Певна увага приділяється майбутнім стандартам моделювання далекого користувача (профіль користувача MPEG-7), стандартам, що забезпечують універсальний доступ до контенту (MPEG-21), а також короткому аналізу механізмів семантичного опису контенту. Ці ідеї та технології мають спростити та покращити майбутній розвиток персоналізованих мультимедійних систем.

Отже, результати нашого дослідження підтверджують ефективність гібридного підходу, який поєднує переваги колаборативної та контентної фільтрації. Порівняно з іншими роботами, наш метод демонструє збалансованість між точністю, охопленням і різноманітністю, що робить його універсальним для широкого кола застосувань. Перспективи подальших досліджень містять інтеграцію реальних даних користувачів і вдосконалення процесу використання обчислювальних ресурсів для розширення масштабованості системи.

Отже, внаслідок виконаної роботи можна сформулювати такі наукову новизну та практичну значущість результатів дослідження.

Наукова новизна отриманих результатів дослідження – розроблено метод підбору персоналізованого контенту, реалізація якого здійснюється через гібридний алгоритм, що поєднує колаборативну фільтрацію із контентною фільтрацією, а також забезпечує високу точність, покриття та різноманітність рекомендацій, даючи можливість ефективно адаптувати систему до динамічних вподобань користувачів і зменшувати проблему "cold start" (холодного старту).

Практична значущість результатів дослідження – розроблений метод підбору персоналізованого контенту, реалізація якого здійснюється через гібридний алгоритм, можна використати для впровадження у реальні соціальні мережі, що сприятиме підвищенню задоволеності користувачів, збільшенню їхньої залученості, створенню конкурентних переваг платформи, а також стане основою для подальшого вдосконалення рекомендаційних систем і розроблення нових методів персоналізації.

Висновки / Conclusion

Визначено найефективніші підходи до підбору персоналізованого контенту для соціальної мережі шляхом поєднання колаборативної та контентної фільтрації, що дало змогу ефективно адаптувати систему до динамічних вподобань користувачів і зменшити проблему холодного старту – "cold start". За результатами проведеного дослідження можна зробити такі основні висновки.

1. Проаналізовано сучасні підходи до персоналізації контенту із використанням нейронних мереж, враховуючи колаборативну і контентну фільтрації, матричну факторизацію та їхні гібридні комбінації. Визначено переваги та обмеження кожного методу, що дало можливість обґрунтувати доцільність застосування гібридного підходу для вирішення завдань персоналізації.
2. Розроблено та впроваджено кілька алгоритмів рекомендацій, які охоплюють ключові методи персоналізації контенту. Проведено їхнє тестування на уніфікованому наборі векторних характеристик мультимедійного контенту (зображень, відео та аудіо), що забезпечило комплексне оцінювання ефективності в різних сценаріях.
3. Реалізовано підготовку даних та створення векторних характеристик для мультимедійного контенту із застосуванням моделей CLIP, I3D та VGGish, що дали можливість врахувати семантичні особливості контенту та підвищити точність рекомендацій.
4. Проведено оцінювання продуктивності алгоритмів за ключовими метриками (Precision@K, Recall@K, NDCG, Coverage, Diversity), що дало можливість визначити гібридний підхід як найкращий серед протестованих методів, із точністю рекомендацій (Precision@10) на рівні 88 % та високою різноманітністю.
5. Здійснено порівняння результатів роботи алгоритмів і встановлено, що гібридний алгоритм ефективно вирі-

шує проблему "cold start", демонструє адаптивність до різноманітних уподобань користувачів і забезпечує оптимальне поєднання точності та різноманітності контенту.

6. Запропоновано рекомендації щодо впровадження гібридного підходу у реальні системи персоналізації контенту, що дає змогу підвищити рівень задоволеності користувачів, їхню залученість, а також забезпечити конкурентні переваги платформи.

References

1. Cheng, W., Yin, G., Dong, Y., Dong, H., & Zhang, W. (2016). Collaborative Filtering Recommendation on Users Interest Sequences. *PLOS ONE*, 11(5), article ID e0155739. <https://doi.org/10.1371/journal.pone.0155739>
2. Fu, W. (2023). AI-News Personalization System Combining Complete Content Characterization and Full Term Interest Portrayal In the Big Data Era. *IEEE Access*, 1. <https://doi.org/10.1109/access.2023.3303479>
3. Hou, L., Chu, D., & Liao, L.-Z. (2023). Convergence of a Fast Hierarchical Alternating Least Squares Algorithm for Nonnegative Matrix Factorization. *IEEE Transactions on Knowledge and Data Engineering*, 1–12. <https://doi.org/10.1109/tkde.2023.3279369>
4. Kim, J., & Jeon, Y. (2022). Personalized User Recommendation based on Various User Behavior in Local Domain. In: *2022 IEEE International Conference on Data Mining Workshops (ICDMW)*. Orlando, FL, USA, 2022, pp. 475–481. <https://doi.org/10.1109/icdmw58026.2022.00069>
5. Liu, H. (2022). Implementation and Effectiveness Evaluation of Four Common Algorithms of Recommendation Systems – User Collaboration Filter, Item-based Collaborative Filtering, Matrix Factorization and Neural Collaborative Filtering. In: *2022 International Conference on Cloud Computing, Big Data Applications and Software Engineering (CBASE)*. Suzhou, China, 2022, pp. 224–227. <https://doi.org/10.1109/cbase57816.2022.00049>
6. Matharu, H., Pasha, Z., Aarif, M., Natrayan, L., Kaliappan, S., & Raj, I. I. (2024). Developing an AI-Driven Personalization Engine for Real-Time Content Marketing in E-commerce Platforms. In: *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*. Kamand, India, 2024, pp. 1–6. <https://doi.org/10.1109/icccnt61001.2024.10725400>
7. Mulyana, R. S., Id Hadiana, A., & Ramadhan, E. (2023). Recommendation System Of Product Sales Ideas For MSMEs Using Content-based Filtering and Collaborative Filtering Methods. In: *2023 International Conference on Computer Science, Information Technology and Engineering (ICCoSITE)*. Jakarta, Indonesia, 2023, 252–256. <https://doi.org/10.1109/icccosite57641.2023.10127844>
8. Pogacnik, M., Tasic, J. F., & Dobravec, S. (2004). Trends in personalized content retrieval and access. *SympoTIC '04. Joint 1st Workshop on Mobile Future & Symposium on Trends In Communications* (IEEE Cat. No.04EX877), Bratislava, Slovakia, 2004, pp. 204–207. <https://doi.org/10.1109/TIC.2004.1409534>
9. Raziperchikolaei, R., & Chung, Y.-J. (2022). A Recommendation System Framework to Generalize AutoRec and Neural Collaborative Filtering. In: *2022 IEEE International Conference on Data Mining Workshops (ICDMW)*. Orlando, FL, USA, 2022, 1162–1167. <https://doi.org/10.1109/icdmw58026.2022.00151>
10. Shrivastava, N., & Gupta, S. (2021). Analysis on Item-Based and User-Based Collaborative Filtering for Movie Recommendation System. In: *2021 5th International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques (ICEECOT)*. Mysuru, India, 2021, 654–656. <https://doi.org/10.1109/iceecot52851.2021.9707955>
11. Vörös, T., Bergeron, S. P., & Berlin, K. (2023). Web Content Filtering Through Knowledge Distillation of Large Language Models. In: *2023 IEEE International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*. Venice, Italy, 2023, 357–361. <https://doi.org/10.1109/wi-iat59888.2023.00058>

12. Zhang, C., & Lu, Y. (2021). Study on artificial intelligence: The state of the art and future prospects. *Journal of Industrial Information Integration*, 23, article ID 100224. <https://doi.org/10.1016/j.jii.2021.100224>
13. Zhang, Z. (2023). Exploration of Hybrid Recommendation Algorithms for Learning Nature Software. In: *2023 2nd International Conference on Data Analytics, Computing and Artificial Intelligence (ICDACAI)*. Zakopane, Poland, 2023, pp. 530–534. <https://doi.org/10.1109/icdacai59742.2023.00106>

B. V. Dutchak, T. O. Korotyeyeva

Lviv Polytechnic National University, Lviv, Ukraine

SELECTION OF PERSONALIZED CONTENT FOR THE SOCIAL NETWORK

The rapid evolution of social networks necessitates the development and implementation of efficient content personalization systems to enhance user engagement and satisfaction. This study analyzed various approaches to content personalization, including collaborative filtering, content-based filtering, matrix factorization, and their hybrid combinations integrating neural networks. State-of-the-art models such as CLIP for images, I3D for videos, and VGGish for audio were applied to generate unified vectorized representations of multimedia data, forming the foundation for algorithm evaluation. The research evaluated the precision, coverage, diversity, and ranking quality of each algorithm. It was established that the hybrid approach, combining collaborative and content-based filtering, achieved the highest recommendation accuracy (Precision@10 = 88 %), along with superior coverage and diversity. The hybrid method effectively addressed the "cold start" problem and demonstrated adaptability to dynamic user preferences. Limitations of standalone methods, including matrix factorization and NCF, were identified, particularly in handling sparse data and maintaining diversity, highlighting the strengths of the hybrid solution. The study introduced strategies for integrating neural networks with traditional algorithms, showcasing their ability to enhance system flexibility and accuracy. Practical implications of the research were presented, providing developers of social networks and streaming platforms with insights to improve user engagement. Further research directions include the integration of real-world user interaction data, the enhancement of hybrid algorithms, and the exploration of advanced models for multimedia content processing. Additionally, computational efficiency was analyzed, with recommendations for reducing training time and resource consumption.

Keywords: content personalization; hybrid algorithms; neural networks; collaborative filtering; content-based filtering.