УДК 004.85:004.89 *http://dx.doi.org/10.30970/vam.2026.36.00000*

РОЗРОБКА РЕКОМЕНДАЦІЙНОЇ СИСТЕМИ ДЛЯ КОЛЕКТИВНИХ РЕКОМЕНДАЦІЙ НА ОСНОВІ МАШИННОГО НАВЧАННЯ З ПІДКРІПЛЕННЯМ

Б. Романюк, О. Пелюшкевич, М. Смичок

Львівський національний університет імені Івана Франка,
вул. Університетська 1, Львів, 79000, Україна,
e-mail: bohdan.romaniuk@lnu.edu.ua,
olga.peliushkevych@lnu.edu.ua, maria.smychok@lnu.edu.ua

Об'єктом дослідження є процес підвищення ефективності надання колективних рекомендацій для груп користувачів у рекомендаційних системах на основі машинного навчання з підкріпленням. Основною проблемою, що розглядається у дослідженні, є покращення точності рекомендацій для груп користувачів, інтереси яких можуть відрізнятися або конфліктувати між собою. Для її вирішення запропоновано модель колективної рекомендаційної системи, що використовує машинне навчання з підкріпленням та механізм урахування впливу кожного окремого користувача на формування представлення інтересів групи. У роботі використано архітектурну модель "Інтегратор-Актор-Критик", реалізовану на основі алгоритму глибинного градієнта детермінованої стратегії, що дає змогу ефективно моделювати послідовний процес прийняття рішень та максимізувати довгострокову винагороду під час формування рекомендацій. Формування рекомендацій здійснюється на основі історичних взаємодій користувачів з елементами та особливостей поведінки групи. Експериментальне дослідження проведено з використанням набору даних MovieLens, а для оцінювання ефективності системи використано метрики Recall@k та NDCG@k. Отримані результати свідчать про потенційну ефективність запропонованої моделі для задач колективних рекомендацій, оскільки вона дає змогу враховувати індивідуальні вподобання учасників групи та покращувати узгодженість сформованих рекомендацій. Практичне значення отриманих результатів полягає у можливості застосування запропонованого підходу у динамічних онлайн-середовищах, зокрема у системах електронної комерції, медіаплатформах, соціальних мережах та новинних ресурсах.

Ключові слова: рекомендаційна система, групові рекомендації, машинне навчання з підкріпленням, модель Актор-Критик.

1. Вступ

У сучасному цифровому світі Інтернет став невід'ємною частиною повсякденного життя людини. Стрімкий розвиток соціальних мереж, електронної комерції та мультимедійних сервісів призвів до значного збільшення обсягів доступної інформації та контенту. У таких умовах виникає проблема інформаційного перевантаження, коли користувачеві стає складно самостійно знайти релевантний контент серед великої кількості доступних варіантів. Для подолання цієї проблеми часто застосовуються рекомендаційні системи, які здатні аналізувати історичну поведінку користувачів та їхні вподобання і пропонувати релевантні об'єкти, наприклад медіаконтент, музичні композиції, новини, товари, тощо. Використання таких систем дає змогу суттєво скоротити час пошуку необхідної інформації, збільшити дохід компаній та підвищити якість користувацького досвіду.

Водночас перспективним напрямом досліджень є застосування машинного навчання з підкріпленням [1], яке дає змогу розглядати процес формування рекомендацій

як послідовність прийняття рішень у динамічному середовищі. У цьому випадку система може поступово адаптуватися до змін уподобань користувачів, враховувати довгострокові наслідки рекомендацій та оптимізувати залученість користувачів.

У ряді практичних застосувань рекомендації повинні формуватися не для окремого користувача, а для групи користувачів. Однак існуючі методи побудови групових рекомендацій такі як MoSAN [2], AGREE [3] та GroupIM [4] характеризуються низкою викликів. По-перше, складно враховувати динамічні зміни вподобань користувачів, оскільки інтереси людей можуть швидко змінюватися з плином часу. По-друге, більшість існуючих моделей оптимізують лише короткострокову винагороду, ігноруючи довгостроковий вплив рекомендацій. По-третє, сучасні групові рекомендаційні системи зазвичай формують рекомендації на основі одного статичного профілю групи та не враховують взаємозв'язки між різними елементами контенту, що може призводити до одноманітності рекомендацій. Тому доцільно враховувати історичність взаємодій для підвищення різноманітності запропонованих елементів.

Для подолання зазначених обмежень у роботі запропоновано колективну рекомендаційну систему на основі машинного навчання з підкріпленням. Запропонована модель містить середовище з одним агентом, який реалізований за допомогою поширеної архітектурної моделі, а його оптимізація здійснюється за допомогою градієнту глибокої детермінованої стратегії.

Практичне значення досліджень у цій галузі полягає в можливості підвищення ефективності сучасних інформаційних сервісів, зокрема платформ потокового медіа для спільного перегляду фільмів і серіалів, систем електронної комерції для сімейних або колективних покупок, туристичних сервісів для планування подорожей групою осіб, а також соціальних мереж і цифрових платформ, де контент споживається або обирається колективно. Використання більш досконалих моделей групових рекомендацій дає змогу покращити релевантність запропонованого контенту, підвищити задоволеність користувачів та сприяти зростанню ефективності цифрових сервісів. Подібні підходи також є корисними в інформаційних системах, у яких дані про окремого користувача є обмеженими або неповними. У таких випадках можливим рішенням є визначення групи користувачів зі схожими вподобаннями та формування рекомендацій на основі колективних інтересів цієї групи.

2. Аналіз літературних даних та постановка проблеми

У роботі [2] запропоновано модель MoSAN (*Medley of Sub-Attention Networks*) для задачі групових рекомендацій, у якій формування рекомендацій базується на врахуванні взаємодій між учасниками групи за допомогою механізму уваги. Результати дослідження свідчать, що такий підхід дає змогу підвищити якість рекомендацій і визначати відносний вплив окремих користувачів у групі. Водночас недоліком моделі є відсутність урахування часової динаміки групових уподобань. Одним із можливих напрямів покращення є поєднання цього підходу з методами, що дають змогу моделювати послідовну зміну інтересів групи з часом.

У роботі [3] запропоновано модель AGREE (*Attentive Group Recommendation*) для групових рекомендацій, у якій групове представлення формується за допомогою механізму уваги з урахуванням взаємодій між користувачами, групами та елементами. Показано, що такий підхід дає змогу підвищити якість рекомендацій і динамічно визначати внесок окремих учасників групи залежно від елемента. Водночас обмеженням моделі є недостатнє врахування часової динаміки групових уподобань і послі-

довного характеру взаємодії з рекомендаційною системою. Одним із можливих напрямів удосконалення є використання методів машинного навчання з підкріпленням, які дають змогу моделювати довгостроковий ефект рекомендацій.

У роботі [4] запропоновано модель GroupIM (*Group Information Maximization*) для формування рекомендацій епізодичним групам з обмеженою історією спільних взаємодій. Показано, що використання максимізації взаємної інформації та адаптивного врахування індивідуальних уподобань дає змогу зменшити вплив розрідженості даних і підвищити якість рекомендацій. Водночас обмеженням моделі є відсутність явного врахування часової динаміки групових уподобань і залежність від налаштування регуляризаційних параметрів. Одним із можливих напрямів покращення є використання методів машинного навчання з підкріпленням, які дають змогу моделювати рекомендаційний процес у динамічному середовищі.

У роботі [5] запропоновано модель DRGR (*Deep Reinforcement Learning based Group Recommender*), у якій задачу групових рекомендацій формалізовано як процес прийняття рішень Маркова та розв'язано за допомогою методів машинного навчання з підкріпленням. Показано, що такий підхід дає змогу враховувати динаміку зміни вподобань і довгостроковий ефект рекомендацій, а також забезпечує кращі результати порівняно з GroupIM. Водночас обмеженнями моделі є спрощена архітектура та залежність від способу формування навчального набору даних. Одним із можливих напрямів удосконалення є використання більш стійких підходів до підготовки даних і розширення архітектури моделі для точнішого врахування групових уподобань.

Отже, аналіз літературних джерел показує, що сучасні підходи до групових рекомендацій пройшли шлях від простих статичних методів об'єднання даних до моделей, які враховують вплив окремих учасників групи, взаємодії між ними та використовують машинне навчання з підкріпленням. Водночас залишаються відкритими питання одночасного врахування динаміки групових рішень, розрідженості даних, індивідуальних історій користувачів і довгострокового впливу рекомендацій. Все це дає підстави стверджувати, що доцільним є проведення дослідження, присвяченого розробці моделі для групових рекомендацій із використанням машинного навчання з підкріпленням, здатного адаптуватися до змін середовища та вподобань користувачів.

3. Мета та задачі дослідження

Метою дослідження є розробка рекомендаційної системи для генерування групових рекомендацій на основі машинного навчання з підкріпленням, яка забезпечує врахування динаміки зміни вподобань користувачів, індивідуальних особливостей членів групи та довгострокового ефекту рекомендацій. Це дасть можливість підвищити точність, адаптивність і релевантність групових рекомендацій у сучасних інформаційних системах. Очікуваним результатом є побудова та експериментальна перевірка моделі, придатної для використання в рекомендаційних системах, де рішення приймаються не окремим користувачем, а групою користувачів.

Для досягнення мети були поставлені наступні задачі:

- здійснити попередню обробку даних для підготовки їх до використання в моделі, описати механізм формування груп користувачів і структуру взаємодій між групами, користувачами та елементами, а також отримати статистичні характеристики сформованого набору даних;

- розробити модель машинного навчання з підкріпленням на основі архітектури Актор-Критик з використанням оптимізатора градієта детермінованої стратегії та методи векторного представлення груп, їх взаємодії із елементами та механізми врахування впливу кожного користувача на групу;
- провести навчання та тестування запропонованої системи у середовищі з використанням експериментального набору даних;
- порівняти ефективність запропонованої моделі із базовим підходом, застосувавши відповідні метрики.

4. Матеріали та методи дослідження

Об'єктом дослідження є процес формування групових рекомендацій у рекомендаційних системах на основі машинного навчання з підкріпленням. Основною гіпотезою дослідження є припущення, що застосування моделі групових рекомендацій, реалізованої на основі архітектури Актор-Критик з використанням нейромережевих представлень стану груп та важливості кожного окремого користувача, дозволить підвищити точність та адаптивність рекомендацій для груп користувачів. Це досягається за рахунок урахування динаміки зміни вподобань, індивідуальних особливостей учасників групи та їхнього спільного впливу на колективне рішення.

У роботі прийнято припущення, що вподобання користувачів і груп можуть бути адекватно відображені у багатовимірному просторі, а історичні взаємодії з елементами системи є репрезентативними для формування подальших рекомендацій. Додатково висувається припущення, що найновіші взаємодії користувачів із системою є більш інформативними для формування рекомендацій, ніж давніші, оскільки вони повніше відображають актуальний стан інтересів як окремих учасників, так і групи в цілому. У межах дослідження використано спрощення, згідно з яким взаємодія групи з елементами системи подається у вигляді єдиного числового вектора, що формується з урахуванням векторних представлень об'єктів, характеристик членів групи та історії взаємодій.

Для аналізу ефективності моделі використано дві основні метрики: $Recall@k$ та $NDCG@k$ [6]. Метрика Recall@k відображає здатність системи знайти всі релевантні елементи серед топ-k запропонованих і задається формулою

$$Recall@k = \frac{\sum_{i=1}^{k} rel(i)}{rel_{count}}. \tag{1}$$

Метрика NDCG@k (нормалізований *Discounted Cumulative Gain*) оцінює якість ранжування рекомендованих елементів, надаючи більшу вагу правильно визначеним елементам, розташованим на вищих позиціях, і обчислюється як

$$NDCG@k = \frac{\sum_{i=1}^{k} \frac{rel(i)}{\log_2(i+1)}}{\sum_{i=1}^{\min(rel_{count},k)} \frac{1}{\log_2(i+1)}}. \tag{2}$$

У формулах (1)-(2) $rel(i) = 1$, якщо i-та рекомендація є релевантною, і $rel(i) = 0$ – інакше, а rel_{count} – загальна кількість релевантних елементів для групи.

Використання зазначених метрик дає змогу здійснити комплексне оцінювання якості рекомендаційної вибірки. Зокрема, $Recall@k$ відображає частку релевантних

елементів, які були виявлені системою, що дає змогу оцінити повноту охоплення релевантного вмісту. Метрика $NDCG@k$ враховує не лише наявність релевантних елементів у рекомендаційному списку, а й порядок їх розташування, що дає змогу оцінити якість ранжування рекомендованих об'єктів. Значення метрик $Recall@k$ та $NDCG@k$ належать до інтервалу $[0, 1]$, де більші значення відповідають вищій якості рекомендацій. Максимальне значення обох метрик дорівнює 1. Для $Recall@k$ це означає, що всі релевантні елементи потрапили до перших k запропонованих, тоді як для $NDCG@k$ значення 1 відповідає ідеальному впорядкуванню релевантних елементів у рекомендованому списку.

Експериментальна частина роботи створена у вигляді консольної аплікації, розробленої мовою програмування Python версії 3.12 (США). Моделі навчання з підкріпленням, функції навчання та валідації реалізовано з використанням бібліотеки PyTorch. Для генерації векторних подань об'єктів і обчислення оціночних метрик було використано додаткові програмні бібліотеки, зокрема scikit-learn версії 1.8.0 та SciPy версії 1.17.1. Для побудови середовища взаємодії застосовано пакет gymnasium версії 1.2.3. Апробацію та валідацію запропонованого підходу здійснено на персональному комп'ютері з використанням апаратного прискорення графічним процесором.

Апаратна конфігурація дослідницького середовища включала восьмиядерний процесор Intel Core i7-9700K, 32 ГБ оперативної пам'яті DDR4 та графічний адаптер NVIDIA GeForce GTX 1650 із 4 ГБ відеопам'яті GDDR5. Як основну операційну систему використано Ubuntu 24.04. Усі обчислювальні експерименти проводилися локально, без залучення хмарних обчислювальних ресурсів.

5. Результати дослідження колективної рекомендаційної системи

5.1. Експериментальний набір даних та його підготовка

Для оцінки ефективності застосування колективного навчання з підкріпленням у середовищі, яке є максимально наближеним до реальних рекомендаційних систем, використано набір даних MovieLens [7]. Цей набір містить інформацію про рейтингові оцінки за п'ятибальною шкалою та інформацію про хронологічний порядок, залишені користувачами однойменного сервісу рекомендації фільмів. Набір даних є загальнодоступним і може бути завантажений з офіційного сервісу. У роботі використано актуальний випуск набору станом на травень 2024 року, що містить близько 32 млн оцінок, отриманих від понад 200 тис. користувачів для приблизно 88 тис. фільмів. Хоча спочатку завантажується повний набір даних, для експериментів із нього формується лише 1000 груп користувачів.

Ключовим етапом підготовки даних є формування взаємозв'язків між користувачами, елементами та групами. Нехай $\mathcal{U} = \{u_1, u_2, \ldots, u_{|\mathcal{U}|}\}$ – множина користувачів, $\mathcal{G} = \{g_1, g_2, \ldots, g_{|\mathcal{G}|}\}$ – множина груп, а $\mathcal{I} = \{i_1, i_2, \ldots, i_{|\mathcal{I}|}\}$ – множина елементів (об'єктів рекомендації) відповідно. Тут u_i позначає окремого користувача, де $i = 1, 2, \ldots, |\mathcal{U}|$; g_j – позначає окрему групу користувачів, де $j = 1, 2, \ldots, |\mathcal{G}|$; а i_k – окремий елемент, який був оцінений або може бути рекомендований, де $k = 1, 2, \ldots, |\mathcal{I}|$. На рис. 1 наведено структуру взаємозв'язків між користувачами та елементами, а також підхід для формування груп на основі цих даних.

Процес базується на тому, що випадковим чином формуються групи, кожна з яких складається з 2–4 користувачів. Для кожної сформованої групи виконується

побудова групових оцінок фільмів на основі індивідуальних оцінок її учасників. Якщо кожен член групи поставив певному фільму оцінку в діапазоні від 4 до 5 балів включно, вважається, що цей фільм позитивно оцінений групою, а його групова оцінка дорівнює 1. Якщо всі учасники групи оцінили фільм, але принаймні половина з оцінок є меншою за 4 бали, вважається, що група негативно оцінила фільм і цей фільм отримує оцінку 0. Нехай $G = \{u_1, u_2, \ldots, u_m\}$ – група користувачів, де $2 \leq m \leq 4$, а $r(u,i)$ – оцінка, поставлена користувачем u фільму i. Тоді групову оцінку $R_G(i)$ можна визначити таким чином:

$$R_G(i) = \begin{cases} 1, & \text{якщо } \forall u \in G: \ r(u,i) \in [4,5], \\ 0, & \text{якщо } \ |\{u \in G : r(u,i) < 4\}| \geq 0.5 \cdot |G|, \\ \varnothing, & \text{в інших випадках.} \end{cases} \tag{3}$$

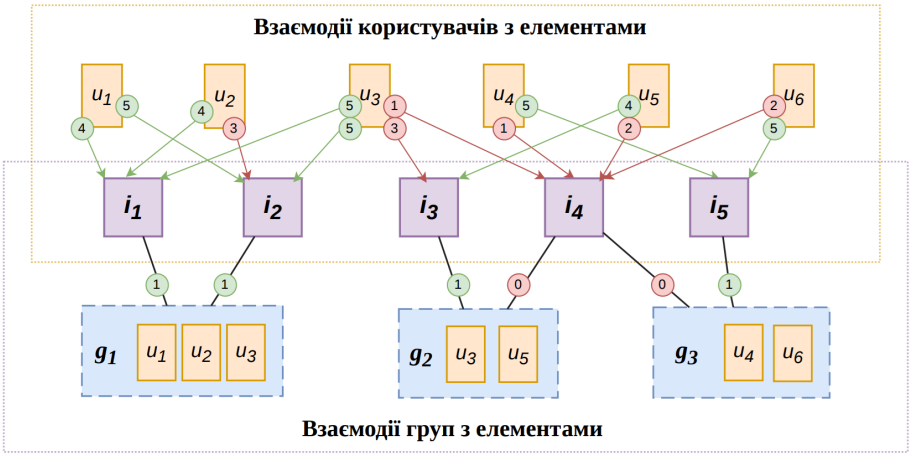

Рис. 1. Схема формування груп та взаємодій користувачів і груп з елементами

В усіх інших випадках, зокрема коли елемент був оцінений менш ніж половиною учасників групи, групова оцінка вважається відсутньою, а такі елементи розглядаються як негативний досвід групової взаємодії, при цьому кількість таких елементів обмежується 100. Для забезпечення достатньої кількості взаємодій у подальшому аналізі до вибірки включаються лише ті групи, для яких сформовано щонайменше 10 групових оцінок. Тобто умова достатньої кількості взаємодій для групи має вигляд: $|\{i \in I \ : \ R_G(i) \in \{0,1\}\}| \ \geq \ 10$. Кількість учасників у групах у межах від 2 до 4 осіб, а також мінімальний поріг у 10 групових оцінок були визначені з урахуванням особливостей набору даних. Подальше збільшення кількості учасників у групі або підвищення мінімального порогу оцінок призвело б до суттєвого зменшення кількості допустимих груп. Це пояснюється тим, що в межах наявних даних імовірність знаходження більшої кількості користувачів зі спільними позитивними взаємодіями щодо однакових елементів є значно нижчою.

Відповідно до визначених правил у результаті формується набір даних, що містить 1000 груп. Детальніші статистичні дані для сформованої вибірки наведено в табл. 1. Варто зазначити, що значення в колонках, які відображають кількість

користувачів та об'єктів взаємодії, відповідають кількості унікальних користувачів і унікальних елементів у межах кожної окремої групи.

Таблиця 1

Статистика сформованого набору груп

Розмір	**К-сть. груп**	**К-сть. корист.**	**К-сть. об'єктів**	**К-сть. оцінок**	**Сер. к-сть. оцінок**	**% позит.**	**% негат.**
2	811	1622	1979	23878	29.44	53.4	46.6
3	164	492	936	4379	26.70	56.9	43.1
4	25	100	388	732	29.28	51.0	49.0
Усього	1000	2214	2093	28989	28.99	53.8	46.2

На цьому етапі сформований набір даних було поділено на три підмножини відповідно до давності взаємодії, що дало змогу зберегти хронологічну послідовність зміни користувацьких і групових уподобань. Зокрема, 70% взаємодій було використано для навчання моделі, а решту 30% – для тестування та валідації. Розподіл здійснювався у хронологічному порядку: взаємодії попередньо впорядковувалися за часом останньої взаємодії, після чого розбивалися на підмножини на основі відповідних відсоткових значень. Такий підхід забезпечує перевірку моделі на новіших взаємодіях порівняно з тими, що використовувалися під час її навчання, і тим самим наближує умови експерименту до реального сценарію функціонування рекомендаційної системи, у межах якого новіші взаємодії мають вищу інформативність порівняно зі старішими.

5.2. Розробка моделі групового машинного навчання з підкріпленням

У межах цього дослідження рекомендаційну систему розглянуто як *агента*, який взаємодіє із середовищем, рекомендуючи об'єкти з метою максимізації винагороди (див. рис.2).

За такого підходу задачу групових рекомендацій доцільно формалізувати як *процес прийняття рішень Маркова* (*Markov Decision Process, MDP*) [8]. Така формалізація дає змогу описати взаємодію рекомендаційної системи із групою користувачів у вигляді послідовності станів, дій і винагород. Формально процес прийняття рішень Маркова задається кортежем з п'яти елементів $(\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma)$, де $\mathcal{S}$ – простір станів, $\mathcal{A}$ – множина дій, $\mathcal{P}$ – імовірність переходів між станами, $\mathcal{R}$ – функція винагороди, а γ – коефіцієнт знижки.

Нехай $\mathcal{U}$, $\mathcal{G}$ та $\mathcal{I}$ позначають множини користувачів, груп та елементів відповідно. Процес прийняття рішень Маркова зображено схематично на рис. 2. В цьому поданні стан $s_t \in \mathcal{S}$ описує поточний стан групи в момент часу t. У даній роботі стан подається у вигляді $s_t = [g, h_t]$, де $g \in \mathcal{G}$ – ідентифікує групу, а h_t – історія переглядів або взаємодій групи. Група g може бути відображена на множину її учасників, тобто $g = \{u_1, u_2, \ldots\}, \quad u_1, u_2, \ldots \in \mathcal{U}$. Історія взаємодій групи подається як $h_t = [i_1, i_2, \ldots, i_N], \quad i_1, i_2, \ldots, i_N \in \mathcal{I}$, де $i_1, i_2, \ldots, i_N$ – об'єкти, з якими група взаємодіяла раніше. Дія $a_t = [i_{t,1}, i_{t,2}, \ldots, i_{t,K}] \in \mathcal{A}$ являє собою список об'єктів, рекомендованих групі рекомендаційною системою, де K – кількість об'єктів у рекомендаційному списку. Після виконання дії a_t у стані s_t рекомендаційна система

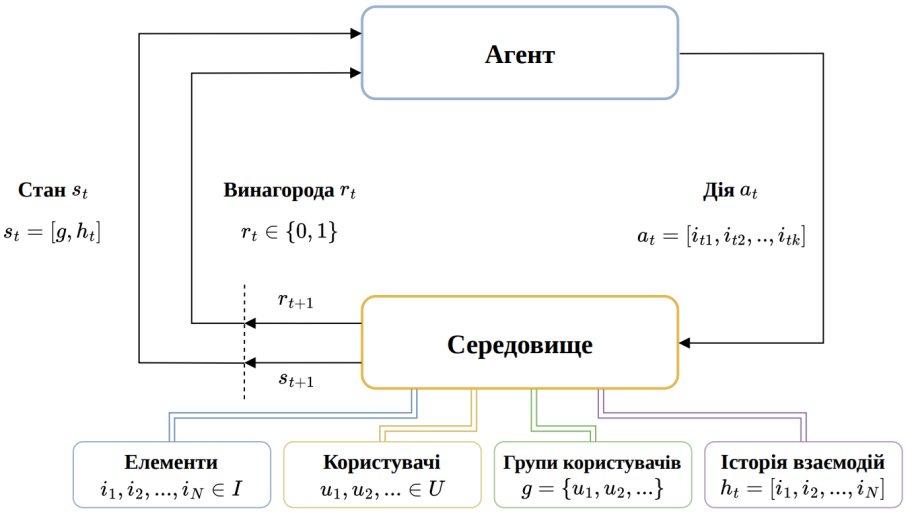

Рис. 2. Процес прийняття рішень Маркова для колективних рекомендацій

отримує винагороду $r_t \in \{0, 1\}$, яка визначається реакцією групи на рекомендований об'єкт. Якщо запропонований об'єкт було позитивно оцінено групою, то $r_t = 1$, інакше $r_t = 0$. Така постановка дає змогу інтерпретувати винагороду як індикатор релевантності рекомендації для групи. Імовірність переходу визначається як $p(s_{t+1} \mid s_t, a_t)$, і задовольняє *властивість Маркова*, відповідно до якої наступний стан залежить лише від поточного стану та виконаної дії. У контексті рекомендаційної системи ця величина характеризує зміну стану середовища в часі після взаємодії групи з рекомендованим об'єктом. Параметр $\gamma \in [0, 1]$ визначає міру врахування майбутньої винагороди під час прийняття поточних рішень. Що ближчим є значення γ до 1, то більшою мірою модель орієнтується на довгостроковий ефект рекомендацій.

Запропонована система побудована на основі одноагентного підходу, у межах якого в середовищі функціонує лише один агент, що навчається на взаємодії з усіма групами та окремими користувачами. Така постановка дає змогу формувати єдину стратегію рекомендацій без розподілу процесу прийняття рішень між кількома агентами. У межах моделі рекомендації можуть генеруватися як для групи користувачів загалом, так і для окремого користувача, який у цьому випадку розглядається як учасник групи. Це забезпечує уніфіковану схему подання вхідних даних, навчання та формування рекомендацій. Водночас ідеї щодо побудови архітектури агента, способу подання стану та врахування групових уподобань ґрунтуються на підходах, запропонованих у роботах [3, 5, 9, 10]. Зокрема, ідею побудови групового середовища, а також механізми врахування впливу окремого користувача на групу, запозичено з праць [5] та [3] відповідно. Архітектурні рішення для навчання моделей і алгоритм навчання, адаптований з урахуванням групової задачі рекомендації, ґрунтуються на підходах, запропонованих у роботах [9] та [10].

Для побудови агента у запропонованій системі використано архітектуру, яка поєднує дві нейронні мережі *Актор* і *Критик* та додаткову мережу для подання стану, яка називається *Інтегратор*. Така структура дає змогу враховувати як

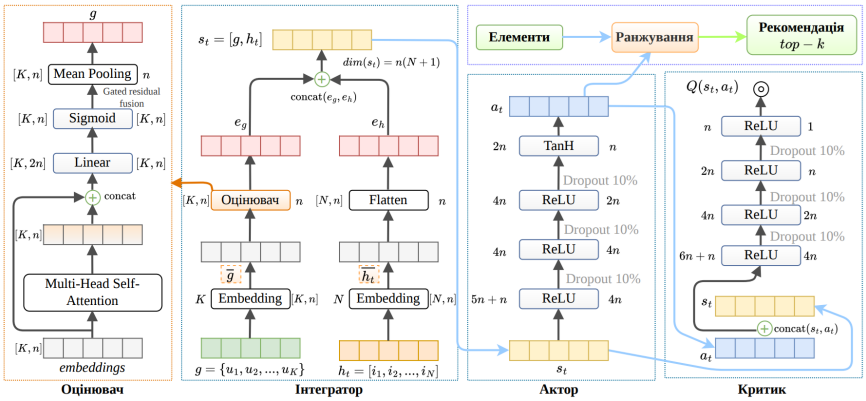

Рис. 3. Архітектура запропонованої колективної рекомендаційної системи

поточні вподобання групи, так і історію її взаємодії з елементами в процесі формування рекомендацій. Архітектуру зазначених мереж наведено на рис. 3. Ліворуч зображено мережу *Інтегратор* та підмодуль *Оцінювач*, які формують узагальнене представлення групи та історії її взаємодії з елементами, у центральній частині – мережу *Актор*, призначену для формування рекомендації, а праворуч – мережу *Критик*, що виконує оцінювання дій актора. На рис. 3 кількість вхідних ознак для повнозв'язних шарів або компонентів архітектури вказано зліва, в той час як кількість вихідних ознак справа.

Мережа Інтегратор (подання стану) приймає на вхід групу користувачів $g = \{u_1, u_2, \ldots, u_K\}$ та історію взаємодій цієї групи з елементами $h_t = [i_1, i_2, \ldots, i_N]$ і формує векторне подання стану s_t. У межах мережі користувачі та елементи історії обробляються окремо за допомогою двох шарів *Embedding* для генерування векторного представлення. На першому етапі кожен користувач $u_j \in g$ перетворюється на вектор ознак фіксованої розмірності n. У результаті формується матриця користувацьких ознак розмірності $K \times n$, у якій кожен рядок відповідає окремому учаснику групи. Отримана матриця надходить до підмодуля *Оцінювач*. На відміну від підходів, де для кожного користувача обчислюється окрема скалярна вага, у запропонованій архітектурі використовується механізм *самоуваги* (*multi-head self-attention*), який дає змогу кожному користувацькому поданню враховувати інформацію про всіх інших членів групи. У результаті цього формується оновлена матриця користувацьких представлень тієї самої розмірності $K \times n$, у якій кожен вектор уже містить контекстну інформацію про взаємозв'язки між учасниками групи. Далі початкові та оновлені представлення об'єднуються, утворюючи матрицю розмірності $K \times 2n$, яка подається на лінійний шар із сигмоїдальною активацією. Цей блок формує коефіцієнти керування розмірності $K \times n$, що визначають, у якій пропорції для кожного користувача слід поєднати початкове та оновлене подання. Після цього виконується кероване злиття ознак, а отримані користувацькі представлення усереднюються за всіма учасниками групи. Таким чином формується єдине групове подання $e_g \in \mathbb{R}^n$. Історія взаємодії групи з елементами обробляється окремо. Для цього використовуються останні N елементів взаємодії, кожен з яких також проходить через власний *Embedding*-шар. У результаті формується багатовимірне подання розмірності $N \times n$. Далі застосовується шар *Flatten*, який

виконує перетворення багатовимірного масиву ознак в один довгий одновимірний вектор. Результатом цього етапу є вектор $e_h \in \mathbb{R}^{N \cdot n}$. Підсумований стан групи в момент часу t формується шляхом конкатенації вектора групового подання та вектора історії взаємодій: $s_t = \text{concat}(e_g, e_h)$. Отже, результатом роботи *Інтегратора* є узагальнене векторне представлення стану $s_t \in \mathbb{R}^{n+N \cdot n}$, яке одночасно враховує групову поведінку користувачів та послідовність їхніх попередніх взаємодій з елементами.

Мережа Актор, яку можна інтерпретувати як мережу поведінкової стратегії, приймає на вхід стан групи s_t, сформований мережею подання стану. Для кожної групи вектор стану послідовно обробляється кількома повнозв'язними шарами з функцією активації *ReLU* (*Rectified Linear Unit*), що дає змогу моделі виявляти нелінійні залежності між характеристиками групи та її історією взаємодій. Після кожного такого шару застосовується механізм *Dropout*, призначений для зменшення ризику перенавчання та підвищення узагальнювальної здатності моделі. На завершальному етапі обробки вихід проміжних шарів надходить до повнозв'язного шару з гіперболічною тангенсною функцією активації. Результатом роботи нейромережі є наближена дія a_t, яка розглядається як вектор із n ознак, що задає подання поточних групових уподобань і використовується для ранжування елементів. Формування рекомендацій здійснюється шляхом обчислення оцінок релевантності для всіх елементів набору даних на основі порівняння вектора дії a_t з їхніми векторними поданнями. Після цього елементи впорядковуються відповідно до отриманих оцінок, а групі пропонується список із топ-k елементів, які мають найвищі значення.

Мережа Критик, яка розглядається як мережа оцінювання, подано у правій частині рис. 3. Її призначення полягає в апроксимації функції винагороди дії $Q(s_t, a_t)$, що залежить від поточного стану групи s_t та дії a_t, сформованої мережею *Актор*. На відміну від мережі *Актор*, яка генерує наближене подання рекомендації, мережа *Критик* оцінює очікувану користь цієї дії з погляду майбутньої винагороди. На вхід мережі *Критик* подаються вектор стану групи s_t, сформований мережею подання стану, а також дія a_t, згенерована актором. На основі конкатенації та аналізу цих вхідних даних мережа обчислює Q-значення, представлене скалярною величиною, яка характеризує якість вибраної дії в заданому стані. У контексті групових рекомендацій це значення відображає, наскільки сформована рекомендація узгоджується з груповими вподобаннями та сприяє максимізації очікуваної винагороди. Оцінки, отримані від мережі *Критик*, використовуються під час оновлення параметрів мережі *Актор* з метою покращення якості сформованої наближеної рекомендації. Таким чином, процес навчання спрямований на вибір таких дій, для яких отримане значення функції $Q(s_t, a_t)$ є максимальним.

5.3. Процедура навчання

Навчання запропонованої моделі для задачі групових рекомендацій здійснюється з використанням алгоритму *градієнту глибинної детермінованої стратегії* (DDPG) [11]. Застосування цього підходу дає змогу поєднати переваги машинного навчання з підкріпленням і глибинних нейронних мереж для побудови адаптивної рекомендаційної системи, орієнтованої на узгодження вподобань усіх учасників групи. На відміну від класичних підходів до рекомендації, у яких рішення приймається для окремого користувача, у даному випадку агент навчається формувати такі рекомендації, що максимізують довгострокову корисність для групи загалом. Це означає, що під час навчання враховуються не лише поточні переваги групи, а й

історія попередніх взаємодій та очікуваний ефект майбутніх рекомендацій.

Алгоритм DDPG передбачає одночасне навчання двох груп нейромереж – основних мереж *Актор* та *Критик*, а також використання відповідних цільових мереж *Актор* та *Критик*, що забезпечують стабільність навчання. Основна мережа Актор позначається як $\pi_\theta(s_t)$, де θ – вектор її параметрів, а основна мережа Критик – як $Q_\phi(s_t, a_t)$, параметризована вектором ϕ. Відповідно цільові мережі позначаються як $\pi_{\theta'}(s_t)$ – цільовий *Актор* та $Q_{\phi'}(s_t, a_t)$ – цільовий *Критик*. Реалізацію алгоритму адаптовано з роботи [10], але для випадку колективних рекомендацій.

У контексті групових рекомендацій дія a_t не інтерпретується як безпосередній вибір одного елемента, а розглядається як неперервне векторне подання поточної рекомендаційної стратегії. На його основі обчислюються оцінки релевантності елементів з використанням косинуса подібності, після чого формується список рекомендацій для групи. Отриманий результат використовується для обчислення винагороди r_t, яка характеризує якість сформованого списку рекомендацій з погляду групової задоволеності, узгодженості інтересів учасників групи або іншої обраної цільової метрики.

Для підвищення стабільності навчання використовується буфер досвіду D, у якому зберігаються переходи вигляду (s_t, a_t, r_t, s_{t+1}). Під час оновлення параметрів мережі *Критик* мінімізується середньоквадратична похибка між передбаченим значенням Q-функції та цільовим значенням, обчисленим за рівнянням Беллмана:

$$L_{\text{critic}}(\phi) = \mathbb{E}_{(s_t,a_t,r_t,s_{t+1})\sim D}\left[\left(Q_\phi(s_t, a_t) - y_t\right)^2\right], \tag{4}$$

де цільове значення y_t визначається як $y_t = r_t + \gamma Q_{\phi'}\left(s_{t+1}, \pi_{\theta'}(s_{t+1})\right)$. У формулі (4) r_t позначає миттєву винагороду, $\gamma \in (0, 1]$ – коефіцієнт знижки, а математичне сподівання $\mathbb{E}$ обчислюється за вибіркою переходів із буфера досвіду D. Використання цільових мереж $\pi_{\theta'}$ та $Q_{\phi'}$ дає змогу зменшити нестабільність навчання, оскільки цільові значення оновлюються поступово, а не змінюються різко на кожній ітерації.

Оновлення параметрів мережі *Актор* здійснюється на основі оцінок, отриманих від мережі *Критик*. Метою є освоєння такої стратегії, яка для кожного стану групи генерує дії з максимально можливим значенням функції $Q(s_t, a_t)$. У межах DDPG градієнт цільової функції для мережі *Актор* апроксимується градієнтом:

$$\nabla_\theta J(\theta) \approx \mathbb{E}_{s_t\sim D}\left[\nabla_a Q_\phi(s_t, a)\big|_{a=\pi_\theta(s_t)} \cdot \nabla_\theta \pi_\theta(s_t)\right]. \tag{5}$$

На практиці це відповідає мінімізації функції втрат мережі *Актор*, яку можна подати у вигляді

$$L_{\text{act or}}(\theta) = -\mathbb{E}_{s_t\sim D}\left[Q_\phi(s_t, \pi_\theta(s_t))\right]. \tag{6}$$

Таким чином, мережа *Актор* навчається формувати такі дії, які, за оцінкою мережі *Критик*, є найбільш доцільними для поточного стану групи та забезпечують максимізацію очікуваної сукупної винагороди. При цьому, підхід для обчислення невідомих винагород r_t для елементів, з якими користувач раніше не взаємодіяв, було запозичено з роботи [5].

Параметри цільових мереж, у свою чергу, оновлюються за механізмом м'якого оновлення, що дає змогу уникнути різких змін цільових значень:

$$\theta' \leftarrow \tau\theta + (1-\tau)\theta', \qquad \phi' \leftarrow \tau\phi + (1-\tau)\phi', \tag{7}$$

де $\tau \in (0, 1)$ – малий коефіцієнт оновлення.

Для векторного подання окремих груп, користувачів та історій взаємодії використано вектори розмірності $n = 64$. Зазначена розмірність є достатньою для репрезентації ключових властивостей відповідних об'єктів, водночас зберігаючи обчислювальну ефективність моделі та забезпечуючи швидше навчання нейронної мережі порівняно з більшими значеннями розмірності, зокрема 128, 256 або 512. Для історії взаємодій беруться останні $k' = 10$ взаємодій користувачів із системою. Параметр τ у формулі (7) встановлено $\tau = 1 \cdot 10^{-3}$. Коефіцієнт знижки встановлено на рівні $\gamma = 0.95$.

Оптимізація всіх трьох мереж – подання стану, *Актор* та *Критик* – здійснюється з використанням оптимізатора *Ranger*, що поєднує переваги алгоритмів *Rectified Adam* (*RAdam*) і *Lookahead* для забезпечення стабільнішої та швидшої збіжності в процесі навчання. Для всіх трьох оптимізаторів швидкість навчання встановлено на рівні $1 \cdot 10^{-3}$, а коефіцієнт регуляризації – на рівні $1 \cdot 10^{-6}$. З метою зниження ймовірності перенавчання в моделі (згідно рис. 3) також використовується механізм *Dropout* з імовірністю 0.1, який передбачає випадкове виключення 10% окремих нейронів під час навчання. Навчання моделі здійснювалося протягом 70 епох, оскільки подальше збільшення кількості епох не приводило до суттєвого покращення значень метрик, що обчислювалися кожні 10 епох. Отже, процедура навчання полягає в ітеративному чергуванні трьох основних етапів: формування дії для поточного стану групи, оцінювання якості цієї дії мережею *Критик* та коригування параметрів мереж *Актор* і *Критик* на основі накопиченого досвіду. Така схема дає змогу поступово наближати стратегію рекомендацій до оптимальної та формувати списки елементів, які найбільшою мірою відповідають груповим уподобанням.

5.4. Оцінка метрик

Порівняльний аналіз запропонованого підходу доцільно здійснювати з моделлю, поданою в роботі [5], оскільки обидва підходи використовують машинне навчання з підкріпленням та архітектуру *Інтегратор–Актор–Критик*. Такий вибір базової моделі дає змогу коректно зіставити ефективність запропонованої архітектури з близьким за концепцією підходом, що належить до того самого класу задач.

Для оцінювання якості рекомендацій використано метрики *Recall@k* та *NDCG@k*. Зазначені метрики обчислювалися як для рекомендацій на рівні окремого користувача, так і для рекомендацій на рівні групи, що дає змогу оцінити модель з погляду точності добору елементів для індивідуального користувача та з погляду її здатності формувати узгоджені рекомендації для групи загалом.

Для формування списків рекомендацій і проведення порівняльного аналізу розглядалися значення $k \in \{5, 10, 15, 20\}$ для кількості результуючих наборів. Вибір таких значень зумовлений тим, що вони відповідають найбільш поширеним розмірам списків рекомендацій та дають змогу оцінити якість моделі як для невеликих, так і для більших наборів результатів. Такий підхід забезпечує порівняння моделей з погляду релевантності рекомендацій, оскільки відображає зміну якості рекомендацій залежно від кількості елементів, запропонованих користувачеві або групі.

У табл. 2 наведено значення використаних метрик для моделі [5] та запропонованого підходу. Усі значення належать до інтервалу від 0 до 1, причому більші значення відповідають кращій релевантності рекомендацій. Найкращі результати виділено жирним шрифтом. За отриманими результатами, запропонована модель демонструє вищу ефективність як для індивідуальних, так і для групових рекомендацій.

Таблиця 2

Результати порівняння моделей на рівні користувачів та груп

k	Модель	Користувачі		Групи	
		Recall@k	NDCG@k	Recall@k	NDCG@k
5	DRGR	0.0868	0.0543	0.1253	0.0780
5	Запроп. модель	**0.1435**	**0.0907**	**0.2318**	**0.1444**
10	DRGR	0.1469	0.0735	0.1871	0.0978
10	Запроп. модель	**0.2138**	**0.1126**	**0.3193**	**0.1717**
15	DRGR	0.2008	0.0878	0.2483	0.1140
15	Запроп. модель	**0.2587**	**0.1245**	**0.3812**	**0.1884**
20	DRGR	0.2502	0.0994	0.2826	0.1221
20	Запроп. модель	**0.3042**	**0.1355**	**0.4253**	**0.1978**

6. Обговорення результатів дослідження колективної рекомендаційної системи

Результати проведеного дослідження демонструють доцільність використання машинного навчання з підкріпленням у задачі групових рекомендацій. Переваги запропонованого підходу зумовлені тим, що модель враховує не лише поточний стан групи, а й історію попередніх взаємодій, що дає змогу формувати рекомендації з урахуванням динаміки групових уподобань. Додатковою перевагою є використання архітектури *Інтегратор–Актор–Критик*, у межах якої мережа *Актор* генерує рекомендаційну стратегію, а мережа *Критик* оцінює її якість з погляду очікуваної винагороди. Такий підхід забезпечує орієнтацію не лише на миттєву релевантність, а й на довгострокову корисність рекомендацій для групи загалом.

За результатами порівняння, наведеними в табл. 2, обрана архітектура забезпечує кращі результати як на рівні окремого користувача, так і на рівні групи. Перевага запропонованої моделі простежується за основними метриками, зокрема *Recall@k* та *NDCG@k*, для розглянутих значень k. Це свідчить про те, що запропонована система не лише точніше виявляє релевантні елементи, а й формує якісніше ранжування. Отримані результати підтверджують ефективність запропонованого підходу, зокрема використання окремої мережі подання стану, механізму врахування впливу учасників групи та історії взаємодій.

Порівняно з моделлю [2], запропонована система демонструє кращі результати завдяки більш повному врахуванню послідовного характеру рекомендаційного процесу. Якщо у [2] основний акцент зроблено на моделюванні взаємодії між учасниками групи та визначенні їхнього відносного внеску у формування групового представлення, то в межах запропонованого підходу додатково оптимізується стратегія вибору рекомендацій з урахуванням довгострокової винагороди. Саме це дає змогу підвищити якість рекомендацій у ситуаціях, коли групові вподобання змінюються в часі або залежать від попередньо рекомендованих елементів.

У порівнянні з роботою [3], яка зосереджена на агрегації індивідуальних уподобань у межах статичного групового представлення, запропонована модель враховує послідовний характер формування рекомендацій. Це дає змогу пов'язувати кожну

наступну рекомендаційну дію з поточним станом групи та історією її попередніх взаємодій, що, своєю чергою, забезпечує кращу адаптацію до змін інтересів у часі.

Порівняно з моделлю [4], запропонований підхід також має переваги, оскільки поєднує побудову узагальненого групового подання зі стратегією послідовної оптимізації рекомендацій. На відміну від підходів, у яких основна увага зосереджена на побудові групового представлення або моделюванні спільних ознак учасників, запропонована архітектура безпосередньо навчається обирати дії, що максимізують очікувану корисність рекомендацій. Це дає змогу не лише краще враховувати структуру групи, а й підвищувати узгодженість між сформованим списком рекомендацій і поточними груповими потребами.

Найбільш показовим є порівняння з роботою [5], оскільки обидва підходи ґрунтуються на використанні машинного навчання з підкріпленням та архітектури *Актор–Критик*. Попри спільну концептуальну основу, запропонована система забезпечує вищі результати. Це може бути зумовлено вдосконаленою процедурою формування стану групи, застосуванням механізму уваги для оцінювання внеску окремих користувачів, а також явним урахуванням історії негативних взаємодій у векторному поданні стану замість випадково обраних даних. У сукупності це створює більш інформативний опис поточної ситуації, на основі якого мережа *Актор* формує точніші рекомендаційні дії, а мережа *Критик* ефективніше оцінює їхню якість.

Отримані результати дають підстави стверджувати, що запропонована модель є перспективним напрямом побудови інтелектуальних систем групових рекомендацій.

7. Висновки

У рамках дослідження було здійснено попередню обробку даних для їх підготовки до використання в моделі, описано механізм формування груп користувачів і структуру взаємодій між групами, користувачами та елементами, а також отримано статистичні характеристики сформованого набору даних.

Розроблено модель машинного навчання з підкріпленням для задачі групових рекомендацій на основі архітектури *Актор–Критик* з використанням алгоритму градієнту детермінованої стратегії. У межах запропонованого підходу реалізовано мережу *Інтегратор*, підмодуль *Оцінювач*, мережу *Актор* і мережу *Критик*, а також побудовано векторне подання груп, користувачів та історії взаємодій і механізм уваги для врахування внеску кожного учасника у формування групового представлення.

Запропоновану модель було навчено та протестовано у середовищі, побудованому на основі створеного експериментального набору даних. У процесі навчання використано цільові мережі та механізм м'якого оновлення параметрів, що забезпечило стабільність оптимізації. Результати тестування підтвердили здатність моделі формувати релевантні рекомендації як для користувачів, так і для груп загалом.

Проведено порівняльний аналіз ефективності натренованої моделі з базовими підходами за метриками $Recall@k$ та $NDCG@k$. За результатами експериментального дослідження встановлено, що запропонований підхід забезпечує вищі показники якості рекомендацій.

Перспективи подальших досліджень пов'язані з вивченням інших способів формування груп, підходів до об'єднання індивідуальних уподобань, а також альтернативних методів урахування впливу окремих користувачів на групове рішення.

Список використаної літератури

1. Chen X. Deep reinforcement learning in recommender systems: A survey and new perspectives / Xiaocong Chen, Lina Yao, Julian McAuley, Guanglin Zhou, Xianzhi Wang // Knowledge-Based Systems. – 2023. – Vol. 264. – DOI: https://doi.org/10.1016/j.knosys.2023.110335
2. Tran L.V. Interact and Decide: Medley of Sub-Attention Networks for Effective Group Recommendation / Lucas Vinh Tran, Tuan-Anh Nguyen Pham, Yi Tay, Yiding Liu, Gao Cong, Xiaoli Li // arXiv:1804.04327. – 2018. – DOI: https://doi.org/10.48550/arXiv.1804.04327
3. Cao D. Attentive Group Recommendation / Da Cao, Xiangnan He, Lianhai Miao, Yahui An, Chao Yang, Richang Hong // Proceedings of the 41st International ACM SIGIR Conference on Research and Development in Information Retrieval. – 2018. – DOI: https://doi.org/10.1145/3209978.3209998
4. Sankar A. GroupIM: A Mutual Information Maximization Framework for Neural Group Recommendation / Aravind Sankar, Yanhong Wu, Yuhang Wu, Wei Zhang, Hao Yang, Hari Sundaram // arXiv:2006.03736. – 2020. – DOI: https://doi.org/10.48550/arXiv.2006.03736
5. Liu Z. Deep Reinforcement Learning based Group Recommender System / Zefang Liu, Shuran Wen, Yinzhu Quan // arXiv:2106.06900. – 2021. – DOI: https://doi.org/10.48550/arXiv.2106.06900
6. Jadon A. A Comprehensive Survey of Evaluation Techniques for Recommendation Systems. / Aryan Jadon, Avinash Patil // Computation of Artificial Intelligence and Machine Learning. – 2024. – P. 281–304. – DOI: https://doi.org/10.48550/arXiv.2312.16015
7. Harper F.M. The MovieLens Datasets. / F. Maxwell Harper, Joseph A. Konstan // ACM Transactions on Interactive Intelligent Systems. – 2015. – Vol. 5 (4). – P. 1–19. – https://doi.org/10.1145/2827872
8. Leon V. Online Reinforcement Learning in Markov Decision Process Using Linear Programming. / Vincent Leon, S. Rasoul Etesami // 62nd IEEE Conference on Decision and Control (CDC). – 1973-1978. – 2023. – DOI: https://doi.org/10.48550/arXiv.2304.00155
9. Zhao X. Deep Reinforcement Learning for List-wise Recommendations / Xiangyu Zhao, Liang Zhang, Long Xia, Zhuoye Ding, Dawei Yin, Jiliang Tang // arXiv:1801.00209. – 2019. – DOI: https://doi.org/10.48550/arXiv.1801.00209
10. Romaniuk B. Development of a multi-agent adaptive recommendation system based on reinforcement learning / Bohdan Romaniuk, Olha Peliushkevych // Eastern-European Journal of Enterprise Technologies. – 2025. – Vol. 5 (2/137). – P. 43–54. – DOI: https://doi.org/10.15587/1729-4061.2025.340491
11. Futuhi E. ETGL-DDPG: A Deep Deterministic Policy Gradient Algorithm for Sparse Reward Continuous Control / Ehsan Futuhi, Shayan Karimi, Chao Gao, Martin Miller // Transactions on Machine Learning Research. – 2025. – DOI: https://doi.org/10.48550/arXiv.2410.05225

Стаття: надійшла до редколегії 02.02.2026
доопрацьована 04.03.2026
прийнята до друку 16.03.2026

DEVELOPMENT OF A GROUP RECOMMENDATION SYSTEM BASED ON REINFORCEMENT LEARNING

B. Romaniuk, O. Peliushkevych, M. Smychok

Ivan Franko National University of Lviv,
1, Universytetska str., 79000, Lviv, Ukraine,
e-mail: bohdan.romaniuk@lnu.edu.ua,
olga.peliushkevych@lnu.edu.ua, maria.smychok@lnu.edu.ua

The object of the study is the process of improving the effectiveness of generating group recommendations for user groups in recommendation systems based on reinforcement learning. The main problem addressed in this study is improving the accuracy of recommendations for groups of users whose interests may differ or conflict with one another. To solve this problem, a group recommender system model is proposed that employs reinforcement learning and a mechanism for taking into account the influence of each individual user on the formation of the group interest representation. The study uses an Integrator-Actor-Critic architectural model implemented on the basis of the Deep Deterministic Policy Gradient algorithm, which makes it possible to effectively model the sequential decision-making process and maximize long-term reward during recommendation generation. Recommendations are generated on the basis of users' historical interactions with items and the characteristics of group behavior. The experimental study was conducted using the MovieLens dataset, and the system performance was evaluated using the Recall@k and NDCG@k metrics. The obtained results indicate the potential effectiveness of the proposed model for group recommendation tasks, as it makes it possible to take into account the individual preferences of group members and improve the consistency of the generated recommendations. The practical significance of the obtained results lies in the possibility of applying the proposed approach in dynamic online environments, in particular in e-commerce systems, media platforms, social networks, and news resources.

Key words: recommender system, group recommendations, reinforcement learning, Actor–Critic model.