

## **ВИКОРИСТАННЯ МЕДІАНИ КЕМЕНІ В АЛГОРИТМІ ФОРМУВАННЯ РЕКОМЕНДАЦІЙ**

**О.С. ЖУРАКОВСЬКА, І.Ю. КОЧУБЕЙ**

**Анотація.** Розглянуто актуальне питання розроблення алгоритмічного забезпечення рекомендаційних систем. Роботу присвячено вирішенню проблеми формування рекомендацій для нових користувачів, в основі якого лежать ідеї переходу від матриці «користувач–предмет» до ранжувань предметів та формування рекомендацій для користувача активного кластера на основі побудови результуючого ранжування, що являє собою медіану Кемені на множині ранжувань. Обґрунтовано вибір медіани Кемені як результуючого ранжування, а також вибір алгоритму її побудови. Для зменшення складності обчислень запропоновано здійснювати агрегацію інформації та використовувати під час формування рекомендації ранжування, побудовані для множини «узагальнених експертів» для даного кластера. Досліджено ефективність розробленого алгоритмічного забезпечення, наведено результати та рекомендації.

**Ключові слова:** рекомендаційна система, узгоджене колективне ранжування, медіана Кемені.

### **ВСТУП**

У зв'язку з поширенням та розвитком електронної комерції, зумовленої розвитком всесвітньої мережі Інтернет, актуальною постала проблема розроблення методів побудови рекомендаційних систем для завдань електронної комерції, метою яких є підвищення лояльності користувачів суб'єкта електронної комерції та максимальне забезпечення потреб користувачів. Однією з найбільш поширених у світі моделей електронної комерції є модель бізнес–споживач (Business–to–Consumer, B2C), до якої належать такі суб'єкти електронної комерції, як інтернет-магазини, електронні дошки оголошень тощо, які мають свої рекомендаційні системи [1].

Аналіз сучасного стану методів побудови рекомендаційних систем показує [2], [3], що актуальною залишається проблема формування рекомендацій для нових користувачів, особливо для нових рекомендаційних систем з огляду на велику розмірність та розрідженість матриці «предмет–користувач». Крім цього, зумовлене збільшенням кількості користувачів та предметів значне зростання розмірності рейтингової матриці «користувач–предмет», у свою чергу, впливає на істотне ускладнення обчислень.

Зазвичай рекомендації часто формуються із застосуванням методів колаборативної фільтрації, асоціативних правил [4, 5, 6].

У випадку формування рекомендацій для нового користувача можна або спочатку прогнозувати початкове наближення профілів оцінок, або сформувати рекомендації, використовуючи узагальнений профіль оцінок користувачів активного кластера. Для реалізації цього підходу необхідно виконати перехід від матриці «користувач–предмет» до ранжувань предметів і перейти до побудови узгодженого колективного ранжування на множині користувачів активного кластера. З огляду на достатньо велику кількість користувачів та предметів, на які повинні бути орієнтовані сучасні рекомендаційні системи (наприклад, кількість користувачів може лежати в інтервалі  $10^3-10^7$ , кількість предметів —  $10^2-10^3$  [7, 8]), актуальною залишається також агрегація використовуваної інформації та підвищення швидкодії алгоритмічного і програмного забезпечення як основи рекомендаційної системи.

### ПОСТАНОВКА ЗАДАЧІ

Є множина користувачів системи, яка в результаті попереднього етапу класифікації розбита на множину кластерів за певними параметрами користувачів [9, 10]. Для існуючих користувачів на основі виставлених ними оцінок предметів сформовано матрицю «користувач–предмет». Розглядається задача формування рекомендацій для нового користувача активного кластера у вигляді ранжування на множині предметів на основі інформації про оцінювання предметів користувачами цього кластера.

Розв’язання поставленої задачі може бути реалізоване виконанням таких етапів:

- побудувати розбиття на множині користувачів активного кластера для агрегації інформації про переваги користувачів;
- для кожної підмножини користувачів з побудованого на попередньому етапі розбиття сформувати ранжування на множині предметів, яке під час формування рекомендацій буде враховуватись як ранжування «узагальненого експерта», який є представником вказаної підмножини користувачів;
- для множини ранжувань, отриманої на попередньому етапі, сформувати результуюче ранжування, яке виражає колективну думку «узагальнених експертів» і може бути використане як рекомендації для нового користувача.

Уведемо ряд позначень:

$r_i = (A_{i_1}, A_{i_2}, \dots, A_{i_n})$  — ранжування на множині альтернатив  $A = \{A_i\}$ ,  $i = \overline{1, n}$ , для якого виконується умова:  $\forall j = \overline{1, n}, \forall k = \overline{1, n} | j < k : A_{i_j} \succ A_{i_k}$ , де  $i_j$  — номер альтернативи, що в ранжуванні  $r_i$  займає позицію  $j$ ; ранжуванню  $r_i$  відповідає відношення, задане матрицею  $P_i = (p_{jk}^i)$ ,  $j = \overline{1, n}, k = \overline{1, n}$ , де

$$p_{jk}^i = \begin{cases} 1, & \text{якщо } A_j \succ A_k, \\ 0, & \text{якщо } A_j \approx A_k, \\ -1, & \text{якщо } A_j \prec A_k; \end{cases} \quad (1)$$

$d(r_1, r_2)$  — відстань між довільними ранжуваннями  $r_1$  і  $r_2$ :

$$d(r_1, r_2) = \frac{1}{2} \sum_{i,j=1}^n |p_{ij}^1 - p_{ij}^2| = \sum_{i < j} |p_{ij}^1 - p_{ij}^2|; \quad (2)$$

$d_{\max}$  — максимально можлива відстань між ранжуваннями:

$$d_{\max} = \max_{i,j} d(r_i, r_j) = n(n-1), \quad (3)$$

де  $n$  — кількість альтернатив;

$d_r$  — сумарна відстань між ранжуванням  $r$  і ранжуваннями користувачів множини  $M$ :

$$d_r = \sum_{i=1}^m d(r, r_i), \quad (4)$$

де  $d(r, r_i)$  — відстань між ранжуваннями  $r$  і  $r_i$ , визначається за формулою (2).

Отже, наведемо математичну постановку задачі.

Задано множину  $M$  — користувачів активного кластера,  $m = |M|$  — кількість користувачів,  $A = \{A_i\}$ ,  $i = \overline{1, n}$  — множину альтернатив (предметів) і матрицю «користувач–предмет»  $T = (t_{ij})$ ,  $i = \overline{1, m}$ ,  $j = \overline{1, n}$ , де  $t_{ij}$  — оцінка користувачем  $i$  альтернативи  $j$ . Кожному користувачу  $k \in M$  відповідає вектор оцінок альтернатив  $v_k = (t_{k1}, t_{k2}, \dots, t_{kn})$ . Необхідно:

1) за множиною векторів оцінок  $\{v_i\}$ ,  $i = \overline{1, m}$  сформуванати множину ранжувань  $\{r_i\}$ ,  $i = \overline{1, m}$  на множині  $A$  для всіх користувачів множини  $M$ ;

2) визначити узгоджене колективне ранжування, розташоване на мінімальній сумарній відстані від ранжувань для всіх користувачів множини  $M$ :

$$r^* = \arg \min_{r \in \Omega} \sum_{i=1}^m d(r, r_i), \quad (5)$$

де  $\Omega$  — множина всіх можливих ранжувань на множині  $A$ ;  $d(r, r_i)$  — відстань між ранжуваннями  $r$  і  $r_i$ , визначається за формулою (2).

## ОБГРУНТУВАННЯ МЕТОДУ РОЗВ'ЯЗАННЯ ЗАДАЧІ

**Вибір медіанного ранжування.** Для пошуку групового консенсусного ранжування ставиться задача знаходження ранжування, яке було б найближчим до всіх індивідуальних ранжувань за деякою введеною мірою відстані. У випадку, коли розв'язок такої задачі визначається співвідношенням (5), він є медіаною Кемені [11, 12]. Оскільки медіана Кемені вважається одним з найбільш коректних результуючих ранжувань, бо вона задовольняє умови 2–5 [11] узгодженого колективного вибору Ерроу (тобто умови універсальності множини допустимих відношень, монотонності, ненав'язуваності та відсут-

ності диктатора), а також задовольняє принцип вибору Кондорсе, не приводячи до парадоксу Кондорсе [11], пропонується на третьому етапі запропонованої вище схеми (побудови консенсусного медіанного ранжування) як результуючого ранжування будувати медіану Кемені.

Задача знаходження медіани Кемені є важкорозв'язуваною комбінаторною задачею і належить до класу NP-повних задач [13, 14, 15, 16]. Існує ряд алгоритмів розв'язання цієї задачі, які базуються на методи гілок і меж, або належать до наближених та евристичних алгоритмів [11, 16, 17, 18, 19].

Оскільки необхідною умовою для алгоритму розв'язання поставленої задачі є можливість розв'язання задач великих розмірностей в режимі реального часу, в запропонованому алгоритмі формування рекомендацій на етапі побудови результуючого ранжування використано евристичний алгоритм, наведений у праці [11]. В алгоритмі для пошуку результуючого ранжування за критерієм мінімізації сумарної відстані до ранжувань множини користувачів  $M$  необхідно сформувати матрицю втрат  $R = (r_{ij})$ ,  $i = \overline{1, n}$ ,  $j = \overline{1, n}$ , елементи якої визначаються співвідношенням

$$r_{ij} = \sum_{k=1}^m d_{ij}(r, r_k), \quad (6)$$

$$\text{де } d_{ij}(r, r_k) = \begin{cases} 0, & \text{якщо } p_{ij}^k = 1, \\ 1, & \text{якщо } p_{ij}^k = 0, \\ 2, & \text{якщо } p_{ij}^k = -1, \end{cases} \quad p_{ij}^k \text{ — елемент матриці відношення, що ві-}$$

дповідає ранжуванню  $r_k$ , визначається за формулою (1).

Тоді для довільного ранжування  $r$  сумарна відстань до всіх ранжувань користувачів множини  $M$  (4) може бути визначена за матрицею втрат:

$$d_r = \sum_{i=1}^n \sum_{j=1}^n r_{ij} p_{ij},$$

де  $r_{ij}$  — елемент матриці втрат (6);  $p_{ij}$  — елемент матриці відношення, що відповідає ранжуванню  $r$  (1).

**Оцінювання отримуваних розв'язків.** Досліджуючи роботу алгоритму побудови медіанного ранжування для задач великої розмірності, доцільно оцінити:

- відхилення значення критерію (4) для розв'язку задачі (5), отриманого на множині ранжувань «узагальнених експертів» (які отримано в результаті агрегації інформації під час розбиття множини  $M$  на підмножини) від значення цього критерію для розв'язку, отриманого на множині ранжувань усіх користувачів з множини  $M$ . Значення відхилення дозволить оцінити доцільність агрегації інформації в межах підмножин розбиття, а також оцінити кількість таких підмножин;
- відносну відстань між вказаними ранжуваннями, використовуючи міру (2) і значення максимально можливої відстані (3). Це дозволить оцінити схожість розв'язків;
- відхилення значення критерію (4) для розв'язку задачі (5), отриманого на множині ранжувань «узагальнених експертів» від значення нижньої

границі критерію (4), яка визначається за матрицею втрат [11] таким співвідношенням:

$$\sum_{i=1}^m d(r, r_i) \leq H_{\min} = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \min \{ r_{ij}, r_{ji} \}. \quad (7)$$

## ДОСЛІДЖЕННЯ

### Опис алгоритму формування рекомендацій новому користувачу на основі побудови колективного ранжування

На вхід алгоритму подається інформація:

- $M$  — множина користувачів кластера ( $m = |M|$ );
- $A = \{A_i\}$ ,  $i = \overline{1, n}$  — множина альтернатив;
- $T = (t_{ij})$ ,  $i = \overline{1, m}$ ,  $j = \overline{1, n}$  — матриця «користувач–предмет»;
- $k$  — кількість підмножин, на які розбивається множина користувачів  $M$  з метою агрегації даних за користувачами. Визначається за результатами проведеного дослідження.

**Крок 1.** Побудувати на множині користувачів  $M$  розбиття  $M = \{M_i\}$ ,  $i = \overline{1, k}$ , де  $k$  — кількість підмножин, що визначається кількістю користувачів у кожній підмножині розбиття.

**Крок 2.** Для кожної підмножини  $M_i$ ,  $i = \overline{1, k}$ , сформувати вектор  $MY_i$  узагальнених оцінок як середнє арифметичне оцінок користувачів даної підмножини.

**Крок 3.** Упорядкувати альтернативи множини  $A$  за спаданням їх оцінок у векторах  $MY_i$  і за цим упорядкуванням сформувати ранжування  $MR_i$ ,  $i = \overline{1, k}$ , на кожній підмножині  $M_i$ . Позначимо отриману множину ранжувань  $MR = \{MR_i\}$ ,  $i = \overline{1, k}$ .

**Крок 4.** Для кожного ранжування  $MR_i$ ,  $i = \overline{1, k}$  побудувати матрицю відношення  $MP_i$  за формулою (1).

**Крок 5.** На основі матриць  $MP_i$ ,  $i = \overline{1, k}$  побудувати відношення  $MP^*$  для узгодженого колективного ранжування за допомогою евристичного алгоритму побудови медіани Кемені.

**Крок 6.** На основі відношення  $MP^*$  сформувати ранжування на множині альтернатив  $A$ , що є рекомендацією для нового користувача кластера.

### Дослідження ефективності алгоритму формування рекомендацій для нового користувача на основі пошуку медіани Кемені

Уведемо позначення:

- алгоритм  $AE$  — евристичний алгоритм побудови медіани Кемені [11];
- $Mr = Mr = \{r_i\}$ ,  $i = \overline{1, m}$  — множина ранжувань усіх користувачів множини  $M$ ;

- $r_1$  — медіана Кемені, знайдена алгоритмом  $AE$  на множині ранжувань  $Mr$  ;
- $r_2$  — медіана Кемені, знайдена алгоритмом  $AE$  на множині ранжувань  $MR$ ;
- $H_{\min}$  — нижнє граничне значення критерію (4), знайдене за співвідношенням (7).
- $d_{r_1}$  — значення критерію (4) для ранжування  $r_1$ , обчислене на множині  $Mr$  ;
- $d_{r_2}$  — значення критерію (4) для ранжування  $r_2$ , обчислене на множині  $Mr$  ;

Проведемо дослідження:

- ефективності алгоритму  $AE$  для знаходження розв'язку на множині  $MR$ ;
- оцінювання кількості користувачів у кожній підмножині  $M_i$ ,  $i = \overline{1, k}$ , яка забезпечує найбільш відповідне співвідношення швидкодії алгоритму і точності результатів.

Для оцінювання ефективності агрегації інформації (використання підмножин  $M_i$  для формування ранжувань для «узагальнених експертів» у побудові результуючого ранжування) будемо порівнювати процес формування ранжування  $r_2$  на множині  $MR$  та формування ранжування  $r_1$  на множині  $Mr$  за швидкодією та наближеністю результатів. Отже, для оцінювання ефективності розробленого алгоритму введемо такі критерії:

- 1) час роботи алгоритму  $AE$  на множині  $MR$ ;
- 2) відношення часу роботи алгоритму  $AE$  для побудови ранжувань  $r_1$  і  $r_2$ ;
- 3) відхилення значень критеріїв  $d_{r_1}$  і  $d_{r_2}$ , обчислених на множині  $Mr$  для ранжувань  $r_1$  і  $r_2$  відповідно:

$$Z = \frac{|d_{r_2} - d_{r_1}|}{d_{r_1}} \cdot 100; \quad (7)$$

відхилення значення критерію  $d_{r_2}$  від  $H_{\min}$  :

$$Q = \frac{|d_{r_2} - H_{\min}|}{H_{\min}} \cdot 100;$$

відносна відстань між ранжуваннями  $r_1$  і  $r_2$  :

$$P = \frac{d(r_1, r_2)}{d_{\max}} \cdot 100, \quad (8)$$

де  $d(r_1, r_2)$  — відстань між ранжуваннями  $r_1$  і  $r_2$  (2);  $d_{\max}$  — максимально можлива відстань між ранжуваннями (3).

Наприклад, відносна відстань між двома ранжуваннями (2,3,1) і (1,2,3) за формулою (8) дорівнюватиме  $P = (2/3) \cdot 100 \approx 67\%$ .

### Обґрунтування розмірності задачі під час проведення досліджень

Аналіз існуючих рекомендаційних систем показав [8], що кількість користувачів може змінюватись у діапазоні  $10^3$ – $10^7$ , а кількість користувачів у кластері — у діапазоні  $10^2$ – $10^4$ . Отже, виконаємо дослідження для кількості користувачів у кластері  $m = 3200$ , що відповідає реальним даним.

### Опис задач дослідження

Одна з сновних задач дослідження — це визначення розміру підмножин  $M_i$  розбиття множини  $M$  для застосування алгоритму  $AE$  на множині  $MR$ . Необхідно оцінити максимальне збільшення розміру групи (підмножини  $M_i$ ), за якого відхилення результатів алгоритму  $AE$  у побудові результуючих ранжувань на множинах  $Mr$  і  $MR$  буде не більшим за 10%. Розмір групи буде змінювати від 50 до  $m/2$  користувачів.

Для практичного застосування алгоритму  $AE$  необхідно оцінити відношення часу роботи алгоритму  $AE$  для побудови результуючих ранжувань на множинах  $Mr$  і  $MR$ .

### РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

1. Відхилення значень критеріїв  $d_{r_1}$  і  $d_{r_2}$  (8), обчислених на множині  $Mr$  для ранжувань  $r_1$  і  $r_2$ , отриманих алгоритмом  $AE$ , а також відхилення значення критерію  $d_{r_2}$  від  $H_{\min}$  (9) (залежно від кількості користувачів у множинах  $M_i$ ,  $i = \overline{1, k}$ ) показано на рис. 1.

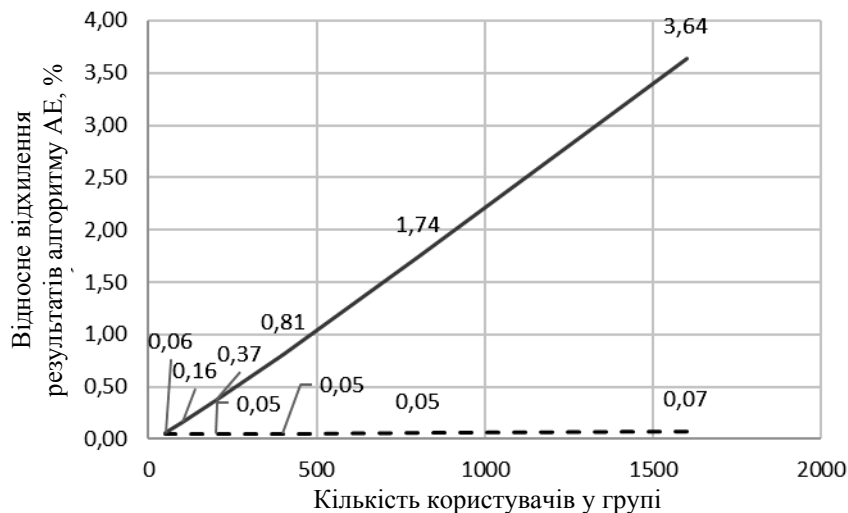


Рис. 1. Відносне відхилення критерію для результату алгоритму  $AE$  на множині  $MR$  (відхилення значень критеріїв  $d_{r_1}$  і  $d_{r_2}$  — пунктирна лінія, відхилення значення критерію  $d_{r_2}$  від  $H_{\min}$  — сувільна лінія)

Як видно з графіка, ранжування  $r_1$  і  $r_2$ , отримані алгоритмом  $AE$  на множинах  $Mr$  і  $MR$  відповідно, за значенням критерію (4) відрізняються дуже мало (менше, ніж на 0,5%). Порівнюючи значення критерію (4) для ранжування  $r_2$  з нижньою границею  $H_{\min}$ , бачимо, що відхилення не перевищує 4%. Якщо розміри множин  $M_i$  від  $\frac{1}{8} \cdot m$  до  $\frac{1}{4} \cdot m$ , то ці результати відрізняються не більше, ніж на 1%. Це дозволяє використовувати алгоритм  $AE$  на множині  $MR$  для побудови результуючого ранжування, яке буде використане у формуванні рекомендацій.

2. Час роботи алгоритму *AE* для побудови результуючого ранжування на множині *MR* залежно від кількості користувачів у множинах  $M_i, i = \overline{1, k}$ , показано на рис. 2.

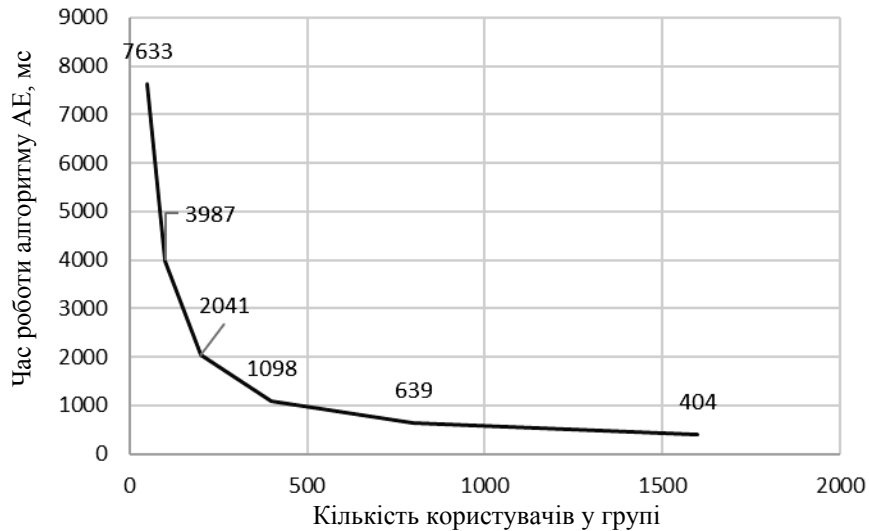


Рис. 2. Залежність часу роботи (мс) алгоритму *AE* для отримання ранжування на множині *MR* від розміру групи

Результати досліджень, подані на рис. 2, дозволяють зробити висновок про можливість застосування алгоритму *AE* для побудови результуючого ранжування на множині *MR* за великих значень кількості користувачів у кластері, що відповідає реальним даним. Для визначеного в попередньому пункті дослідження розміру підмножини  $M_i$  від  $(1/8)t$  до  $(1/4)t$  час роботи алгоритму *AE* на множині *MR* складає менше ніж 2 с.

3. Відношення часу роботи алгоритму *AE* для множини *Mr* і *MR* залежно від кількості користувачів у множинах  $M_i, i = \overline{1, k}$ , подано на рис. 3.

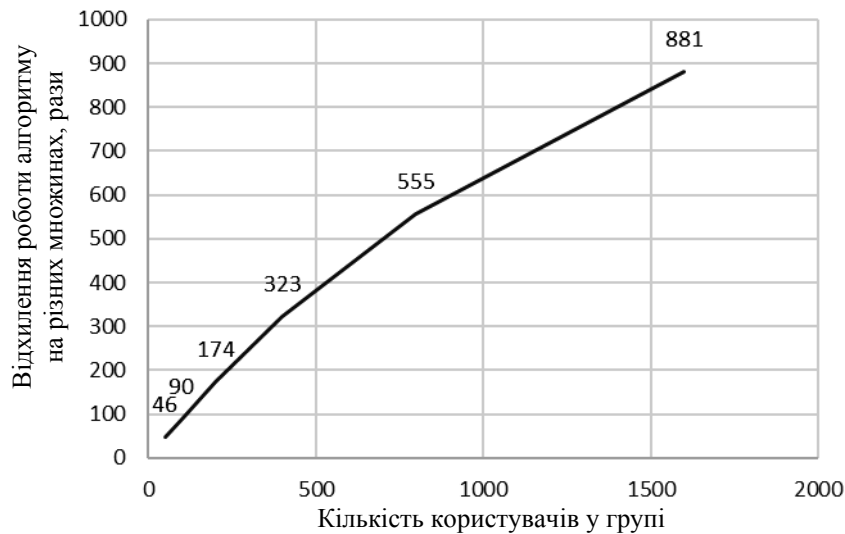


Рис. 3. Відношення часу роботи алгоритму *AE* на множинах *Mr* і *MR* для різних розмірів груп



Порівнюючи час роботи алгоритму *AE* на множинах *Mr* і *MR*, бачимо, що ефективність побудови результуючого ранжування алгоритмом *AE* на множині *MR* від 100 до 350 разів вища за цим критерієм.

4. Відносну відстань між ранжуваннями  $r_1$  і  $r_2$  (10) зображено на рис. 4.

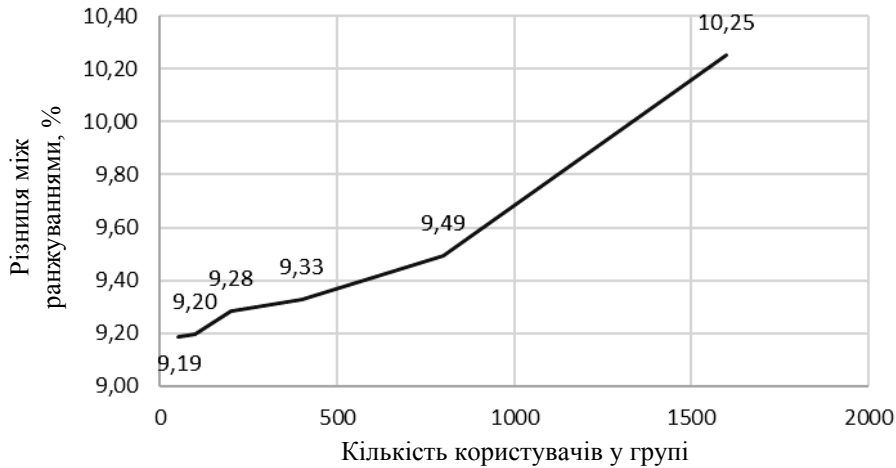


Рис. 4. Відносна відстань між ранжуваннями  $r_1$  і  $r_2$  для різних розмірів груп

Порівнявши відносну відстань між ранжуваннями, побудованими на множині ранжувань користувачів кластера (ранжування  $r_1$ ) і на множині ранжувань «узагальнених експертів» (ранжування  $r_2$ ), бачимо, що вона складає не більше ніж 10% для визначеного в попередніх пунктах дослідження розміру групи. А враховуючи істотно вищу швидкість алгоритму *AE* на множині *MR*, робимо висновок про доцільність практичного використання алгоритму *AE* на множині *MR* під час формування узагальненого ранжування, на основі якого формуються рекомендації для нового користувача.

## ВИСНОВКИ

Проаналізовано проблеми формування рекомендацій в сучасних рекомендаційних системах. Показано, що проблема формування рекомендацій для нового користувача є актуальною. У зв'язку з великою кількістю користувачів та предметів рекомендаційних систем актуальною залишається проблема розроблення алгоритмічного забезпечення, яке дозволяє формувати рекомендації в реальному часі. У роботі запропоновано алгоритм формування рекомендацій, що вирішує проблему холодного старту. В основу покладено ідею використання як колективного ранжування медіани Кемені, на основі якого формуються рекомендації для нового користувача. Для підвищення швидкості алгоритму формування рекомендацій запропоновано використовувати агреговану інформацію про переваги користувачів. Для цього слід розбивати множину користувачів на підмножини, розмірність яких визначається в результаті проведеного дослідження.

Проведено дослідження розробленого алгоритмічного забезпечення. Досліджено відхилення результатів під час побудови результуючого ранжування на множині всіх користувачів та на множині узагальнених користувачів. Показано, що в разі розбиття множини користувачів на підмножини розмірністю від 1/8 до 1/4 від кількості користувачів кластера відносна відхилення результатів не перевищує 1%. Досліджено швидкість запропо-

нованого алгоритму і показано, що з використанням агрегованої інформації за користувачами для формування узагальненого ранжування швидкодія алгоритму підвищується від 100 до 350 разів. Аналіз часу роботи розробленого алгоритму підтверджує можливість його використання на реальних даних у режимі реального часу.

За результатами дослідження можна стверджувати про доцільність використання розробленого алгоритму в рекомендаційних системах для вирішення проблеми формування рекомендацій новому користувачу.

## ЛІТЕРАТУРА

1. J. Lu, D. Wu, M. Mao, W. Wang, and G. Zhang, “Recommender system application developments: A survey”, *Decision Support Systems*, vol. 74, pp. 187–192, 2015. doi:10.1016/j.dss.2015.03.008.
2. C.A. Gomez-Uribe and N. Hunt, “The Netflix Recommender System: Algorithms, Business Value, and Innovation”, *ACM Transactions on Management Information Systems*, vol. 6, no. 4, pp. 1–19, 2015. doi:10.1145/2843948.
3. F. Ricci, L. Rokach, and B. Shapira, “Introduction to Recommender Systems Handbook”, in *Recommender Systems Handbook*, Boston, MA: Springer, 2011, pp. 1–35.
4. P. Melville et al., “Content-Boosted Collaborative Filtering for Improved Recommendations”, in *National Conference on Artificial Intelligence*, Edmonton, Canada, 2016, pp. 187–192.
5. V. Srikar and R. Sudha, “Examining Lists on Twitter to Uncover Relationships Between Following, Membership and Subscription”, in *Proceedings of the 22nd international conference on World Wide Web*, Rio de Janeiro, Brazil, 2013, pp. 673–676.
6. I.Yu. Kochubey and O.S. Zhurakovska, “Modified algorithm of collaborative filtering for forming user recommendations”, *KPI Science News*, vol. 3, pp. 43–49, 2020. doi:10.20535/kpi-sn.2020.3.209842.
7. D.T. Pham, S.S. Dimov, and C.D. Nguyen, “Selection of K in K-means clustering”, *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 219 (1), pp. 103–119, 2005. doi:10.1243/095440605x8298.
8. W. Prinz et al., “PolyLens: A Recommender System for Groups of Users”, in *Proceedings of the Seventh European Conference on Computer-Supported Cooperative Work*, Bonn, Germany, 2001, pp. 199–218. doi:10.1007/0-306-48019-0\_11.
9. A. Shepitsen, J. Gemmell, B. Mobasher, and R. Burke, “Personalized Recommendation in Social Tagging Systems Using Hierarchical Clustering”, in *Proceedings of the ACM Conference on Recommender Systems — RecSys '08*, 2008, pp. 259–266. doi:10.1145/1454008.1454048.
10. C. Gentile et al., “On Context-Dependent Clustering of Bandits”, in *Proceedings of the 34th International Conference on Machine Learning*, 2017, pp. 1253–1262.
11. B. Litvak, *Ekspertnaya informatsiya. Metody polucheniya i analiza* [Expert information. Methods of obtaining and analysis]. Moscow, 2009
12. C. List and C. Puppe, “Judgement aggregation: a survey”, in *Oxford handbook of rational and social choice*, Oxford, 2009, pp. 457–482. doi:10.1007/s11229-011-0025-3.
13. H. Bury and D. Vagner, “Judgement With Ties. Distance-Based Methods”, *New Approaches in Automation and Robotics*, Vienna, 2008, pp. 153–172.
14. O. Hudry, “Complexity of computing median linear orders and variants”, *Electronic Notes in Discrete Mathematics*, vol. 42, pp. 57–64, 2013. doi:10.1016/j.endm.2013.05.146.
15. O. Hudry, “NP-hardness results on the aggregation of linear orders into median orders”, *Annals of Operations Research*, vol. 163, pp. 63–88, 2008. doi:10.1007/s10479-008-0353-y
16. S. Amodio, A. D'Ambrosio, and R. Siciliano, “Accurate algorithms for identifying the median ranking when dealing with weak and partial rankings under the Kemeny

- axiomatic approach”, *Journal of Operational Research*, vol. 249, pp. 667–676, 2015. doi:10.1016/j.ejor.2015.08.048.
17. E. Emond and D. Mason, “A new rank correlation coefficient with application to the consensus ranking problem”, *Journal of Multi-Criteria Decision Analysis*, vol. 11, pp. 17–28, 2002.
  18. W.D. Cook, B. Golany, M. Penn, and T. Raviv, “Creating a consensus ranking of proposals from reviewers partial ordinal rankings”, *Computers & Operations Research*, vol. 43, pp. 954–965, 2007.
  19. A. Alnur and M. Meila, “Experiments with Kemeny Ranking: What Works When?”, *Mathematical Social Sciences*, vol. 64, pp. 28–40, 2012. doi:10.1016/j.mathsocsci.2011.08.008.

Надійшла 09.10.2020

### INFORMATION ON THE ARTICLE

**Oksana S. Zhurakovska**, ORCID: 0000-0002-2804-5556, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine, e-mail: o.zhurakovska@kpi.ua

**Illa Yu. Kochubey**, ORCID: 0000-0003-2520-6577, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine, e-mail: illa.kochubey1@gmail.com

### USE OF KEMENY MEDIAN IN THE ALGORITHM OF FORMING RECOMMENDATION / O.S. Zhurakovska, I.Yu. Kochubey

**Abstract.** The relevant nowadays question of development of the algorithmic support of recommender systems is considered. The article is devoted to the solution of the problem of forming recommendations to new users, which is based on the ideas of transition from the matrix «user-object» to the ranking of objects and the formation of recommendations to the user of the active cluster based on the construction of the resulting ranking, which is a Kemeny median on a set of rankings. The choice of Kemeny median as the resulting ranking and the choice of algorithm for its construction are justified. To reduce the complexity of calculations, it is suggested to perform aggregation of information and to use it in forming of ranking recommendations, which are based on a set of «generalized experts» for this cluster. The efficiency of the developed algorithmic support was studied and the results and recommendations were given.

**Keywords:** recommender system, consensus ranking, Kemeny median.

### ИСПОЛЬЗОВАНИЕ МЕДИАНЫ КЕМЕНИ В АЛГОРИТМЕ ФОРМИРОВАНИЯ РЕКОМЕНДАЦИЙ / О.С. Жураковская, И.Ю. Кочубей

**Аннотация.** Рассмотрен актуальный в настоящее время вопрос разработки алгоритмического обеспечения рекомендательных систем. Работа посвящена решению проблемы формирования рекомендаций новым пользователям, в основе которого лежат идеи перехода от матрицы «пользователь–предмет» к ранжированию предметов и формирования рекомендаций пользователю активного кластера на основе построения результирующего ранжирования, которое представляет собой медиану Кемени на множестве ранжирований. Обоснован выбор медианы Кемени в качестве результирующего ранжирования, а также выбор алгоритма ее построения. Для снижения сложности вычислений предложено осуществлять агрегацию информации и использовать при формировании рекомендации ранжирования, построенные на множестве «обобщенных экспертов» для данного кластера. Исследована эффективность разработанного алгоритмического обеспечения, приведены результаты и рекомендации.

**Ключевые слова:** рекомендательная система, согласованное коллективное ранжирование, медиана Кемени.